The American Economic Review

Cut-H02183-85-P014121

85

ARTICLES

SEPTEMBER 1984

THE AMERICAN ECONOMIC ASSOCIATION

•Published at George Banta Co., Inc., Menasha, Wisconsin. The publication number is ISSN 0002-8282.

•*THE AMERICAN ECONOMIC REVIEW* including four quarterly numbers, the *Proceedings* of the annual meetings, the Directory, and Supplements, is published by the American Economic Association and is sent to all members five times a year: March; May; June; September; December.

Dues for 1984, which include a subscription to both the *American Economic Review* and the *Journal of Economic Literature*, are as follows:

$33.00 for regular members with rank of assistant professor or lower, or with annual income of $15,840, or less;

$39.60 for regular members with rank of associate professor, or with annual income of $15,840 to $26,400;

$46.20 for regular members with rank of full professor, or with annual income above $26,400.

Institutional subscriptions (libraries, institutions, or firms) are $100.00 a year. Individual subscriptions are $65.00 a year. Only subscriptions to both publications will be accepted. Single copies of either journal may be purchased from the Secretary's office.

In countries other than the United States, add $9.20 to cover extra postage.

•Correspondence relating to the Directory, advertising, permission to quote, business matters, subscriptions, membership and changes of address should be sent to the Secretary, C. Elton Hinshaw, 1313 21st Avenue So., Suite 809, Nashville, TN 37212-2786. Change of address must reach the Secretary at least six (6) weeks prior to the month of publication. The Association's publications are mailed second class.

•Second-class postage paid at Nashville, Tennessee and at additional mailing offices, Printed in U.S.A.

•Postmaster: Send address changes to *American Economic Review*, 1313 21st Avenue So., Suite 809, Nashville, TN 37212-2786.

Founded in 1885

Officers

President
CHARLES L. SCHULTZE
The Brookings Institution

President-Elect
CHARLES P. KINDLEBERGER
Massachusetts Institute of Technology

Vice Presidents
ZVI GRILICHES
Harvard University
ROBERT L. HEILBRONER
New School for Social Research

Secretary
C. ELTON HINSHAW
Vanderbilt University

Treasurer
RENDIGS FELS
Vanderbilt University

Managing Editor of The American Economic Review
ROBERT W. CLOWER
University of California-Los Angeles

Managing Editor of The Journal of Economic Literature
MOSES ABRAMOVITZ
Stanford University

Executive Committee

Elected Members of the Executive Committee
ANN F. FRIEDLAENDER
Massachusetts Institute of Technology
JOSEPH E. STIGLITZ
Princeton University
WILLIAM D. NORDHAUS
Yale University
A. MICHAEL SPENCE
Harvard University
VICTOR R. FUCHS
Stanford University
JANET L. NORWOOD
Bureau of Labor Statistics

EX OFFICIO Members
GARDNER ACKLEY
The University of Michigan
W. ARTHUR LEWIS
Princeton University

JAMES M. BUCHANAN

DISTINGUISHED FELLOW

1983

When James Buchanan was studying economics at the University of Chicago, Public Finance was a traditional subdiscipline progressing forward slowly along predetermined lines. In the thirty years since then, there has been a true paradigm shift in Public Finance and the creation of a new, but related field, Public Choice. In both of these intellectual revolutions Buchanan played a leading and, indeed, a dominating role. Today, economics is concerned with public policy, not merely as offering advice as to what is a good policy, but examining in detail government procedures and in analyzing government itself. This is of course the field of Public Choice in which Buchanan has played such a great role.

Classical Public Finance was to a large extent devoted to tax analysis. Buchanan was a major figure in shifting it to the examination of expenditures as well. The immense expansion of research dealing with public goods and externalities bears the imprint of his work. Indeed, in both Public Finance and Public Choice there is now an entire community of scholars whose work deals with subjects that were regarded as either outside of economics or radically different from tradition thirty years ago, and who think of themselves as disciples of Buchanan.

In addition to these major paradigm shifts, Buchanan has made contributions through the entire range of economics. The theory of clubs and his work on the public debt have been particularly influential. There are but few fields where his work is not cited. Altogether, there are few economists whose impact on the profession has been as great.

THE AMERICAN ECONOMIC REVIEW

ROBERT W. CLOWER
Managing Editor

JOHN G. RILEY
Associate Editor

WILMA ST. JOHN
Production Editor

THERESA DE MARIA
Assistant Editor

Board of Editors

GEORGE A. AKERLOF
CLIVE D. BULL
PATRICIA DANZON
MICHAEL R. DARBY
PHILIP E. GRAVES
JACK HIRSHLEIFER
MEIR KOHN
FREDERIC S. MISHKIN
SHERWIN ROSEN
RICHARD SCHMALENSEE
SUSAN WOODWARD
LESLIE YOUNG

• Manuscripts and editorial correspondence relating to the regular quarterly issues of this *Review* and the *Papers and Proceedings* should be addressed to Robert W. Clower, Managing Editor, *AER* Editorial Office, University of California, Los Angeles, CA 90024. Manuscripts should be submitted in triplicate and in acceptable form, and should be no longer than 50 pages of double-spaced typescript. A submission fee must accompany each manuscript: $25 for members; $50 for nonmembers. *Style Instructions* for guidance in preparing manuscripts will be provided upon request to the editor.

• No responsibility for the views expressed by authors in this *Review* is assumed by the editors or the publishers, The American Economic Association.

• Copyright © American Economic Association 1984. All rights reserved.

September 1984

VOLUME 74, NUMBER 4

Articles

p14921

Shorter Papers

Are Workers Paid their Marginal Products?

By ROBERT H. FRANK*

Status is, like Coase's social costs, a reciprocal phenomenon. Given that one person's gain in status can occur only at the expense of a loss in status for others, and that workers are free to choose their co-workers, it follows that the competitive wage structure within a firm must be one in which individual wage differences understate individual differences in marginal products.[1] The purpose of this paper is to examine a variety of empirical evidence that relates to this proposition about the firm's internal wage structure.

The paper is organized as follows. Section I briefly summarizes the theoretical considerations that govern competitive wage determination when status matters to people and firms are viewed as voluntary associations of workers. Section II then confronts the predictions of Section I by examining pay and productivity schedules for a group of sales occupations for which these schedules are relatively easily observed. Section II also examines the relationship between wages and productivity for a sample of university professors, an occupation in which individual productivity differences are, for a variety of obvious reasons, relatively more difficult to measure. All of the evidence examined is consistent with the hypothesis that, within firms, wage rates vary substantially less than do individual productivity values. Section III discusses additional observations and evidence that bear on this same hypothesis. It suggests that the implicit market for status may strongly influence the ways in which firms are organized to carry out the tasks they perform. Section IV concludes by considering the claim that egalitarian internal wage structures arise because of "equity considerations." It argues that the concept of equity appears very closely linked to the concept of status, and suggests a strategy for assigning monetary value to the equity considerations that so often dominate public policy decisions.

I. Freedom of Association, Interdependent Preferences, and the Competitive Wage Structure

Standard neoclassical models in which individual utility functions are independent predict that workers will be paid their marginal products by cost-minimizing firms who purchase their services in competitive labor markets. At a superficial level, at least, this prediction appears sharply at variance with pay schedules observed in practice, which often seem much more egalitarian than would be warranted by productivity considerations alone. Many firms, for example, follow strict pay formulas based on education, experience, and length of tenure with the firm, even when there are apparently very wide differences in productivity among individual workers who are alike with respect to the characteristics specified in the pay formula. Even in circumstances where we do observe pay schedules in which compensation is linked directly to employee output, the slopes of these schedules with respect to productiv-

*Associate Professor of Economics, Cornell University, Ithaca, NY 14853. I thank Randy Lutter for his very able research assistance; Joe Kennedy for his thoughtful collaboration on the material pertaining to the real estate industry; and Associate Dean Geoffrey Chester and Kathleen Platt of Cornell University's College of Arts and Sciences for their extremely generous cooperation in making available the material on Cornell research grants.

[1]See my 1983 paper. Contrary to appearances, this proposition is completely consistent with the observation that, for the labor market as a whole, the distribution of wages is much more dispersed than the underlying distribution of individual ability. Sherwin Rosen (1982) has recently shown, for example, that small differences in ability multiply through layers of organizational hierarchy to create large differences in marginal products. The proposition at issue here concerns how the wage distribution within each firm relates to the corresponding distribution of marginal products, not underlying abilities. Moreover, the proposition I test here does not contradict in any way the traditional claim that wages and marginal products have the same average value within each firm.

ity tend to be substantially less than one (more on these points below).

One possible explanation of the apparently excessive degree of compression observed in many pay schemes is that it is prohibitively costly to measure individual differences in productivity well enough for these differences to be reflected fully in wages.[2] This observation has obvious force when applied to complex team production processes.[3] Yet there are many occupations in which the productive contributions of individual employees are relatively easy to isolate and measure. And as we shall presently see, observed variations in earnings are substantially smaller than observed variations in productive contributions even in occupations such as these.

The possibility that employees are relatively more risk averse than firms cannot satisfactorily explain why wage rates within firms vary so much less than do productivity values.[4] It is arguably the case that individual productivity values are known with relatively less certainty early in a worker's career than later, and that young workers might therefore wish to hedge their bets by accepting employment under a pay scheme more egalitarian in nature than would be called for by subsequently revealed information. Once information on individual productivity values begins to accumulate, however, binding long-term labor contracts would be needed to prevent new firms from bidding successfully for any employee whose wage in an existing firm did not fully reflect his or her marginal product. Yet binding contracts of this sort do not appear to exist in practice. Employees may also prefer gently sloped earnings functions as a means of smoothing their earnings streams when there is uncertainty in the amount they will produce during each period. But this objective can be accomplished in a variety of ways that do not produce unearned benefits for less-productive employees at the expense of their more-productive counterparts. Many automobile dealers, for example, provide salespersons with a steady "draw," which is adjusted periodically so as to remain in balance with commissions earnings over the long run. Important as the desire for a smooth earnings stream may be in some instances, we shall see that it cannot account for the extent to which earnings differences understate productivity differences in the occupations examined below.

The principal-agent problem has also been discussed as a source of compression of the internal earnings distributions of firms.[5] Risk aversion on the part of the agent is what drives the result in principal-agent models that wage rates under optimal labor contracts do not vary one-for-one with what the agent produces. But if risk aversion by itself cannot account for the degree of wage compression we observe in practice (as I shall argue below), the additional complication that output depends on unobservable effort levels cannot do so either. On the contrary, the more strongly a worker's output depends on his level of effort, the more closely linked his wage should be to the value of what he produces.[6]

Numerous writers have suggested that "social" forces may act to flatten the wage vs. marginal product schedule both within and across firms.[7] John Pencavel, for example, suggests that pure piece-rate pay schemes would lead to such widespread pay differences between workers as to "induce discontent among employees, ... followed by uncooperative and unaccommodating work behavior" (1977, p. 239). It is perhaps an understatement to say that economists have

[2]See, for example, Armen Alchian and Harold Demsetz (1972). See also Edward Lazear and Rosen (1981), and Milton Harris and Artur Raviv (1978).

[3]See Theodore Groves (1973).

[4]This possibility has been suggested by Joseph Stiglitz (1975), among others.

[5]See Steven Shavell (1979).

[6]Descriptions of internal labor markets have also been used to account for internally egalitarian wage structures (see, for example, Peter Doeringer and Michael Piore, 1971, and Lester Thurow, 1975). These accounts, which turn on the gains from making major investments in firm-specific human capital, seem plausible enough in many occupations. but firm-specific human capital plays only a minimal role in the particular occupations I shall examine below.

[7]See J. R. Hicks (1955), Paul Streeten (1962), John Burton and John Addison (1977), Daniel Hamermesh (1975), and John Dunlop (1957).

generally reacted coolly to suggestions such as these, for they appear to ignore the obvious objection that any workers not paid the full value of their marginal products by one firm could be successfully bid away from their current positions by other firms that are willing to do so. Yet, for all of its apparent force, this objection loses sight of the fact that when wage schedules are less steep than the standard textbook wage schedule, there results a clear, positive relationship between a worker's status in the income hierarchy of his firm and the extent to which his wage understates his marginal product. The most highly paid worker in a firm with a pay schedule like *AA* in Figure 1, for example, earns ΔW_1 less than his marginal product, MP_1, whereas the worker who occupies the bottom slot in the income hierarchy of the same firm earns ΔW_2 more than his marginal product, MP_2. Abundant evidence exists in support of the claim that individuals care a great deal about what positions they occupy in the income hierarchies of the groups to which they belong.[8] Elsewhere I have shown that if no compensating variations for within-group status are present and if production takes place under constant returns to scale, then the equilibrium firms must consist of workers whose marginal products are identical.[9] In the absence of a compensating wage differential, what reason would the worker with the marginal product MP_2 in Figure 1 have for enduring a low-status position amongst his coworkers? He could clearly do better by joining some other firm composed only of workers whose productivities and wages were identical to his own.

Now, the refusal by the less-productive worker to associate with the more-productive worker when wages are equal to marginal products leads to a gain in welfare for the less-productive worker, but creates a loss for his more-productive counterpart, who would otherwise have been able to enjoy a position of relatively high status within his work group. As with Coase's question of whether a firm should pollute, the question of whether

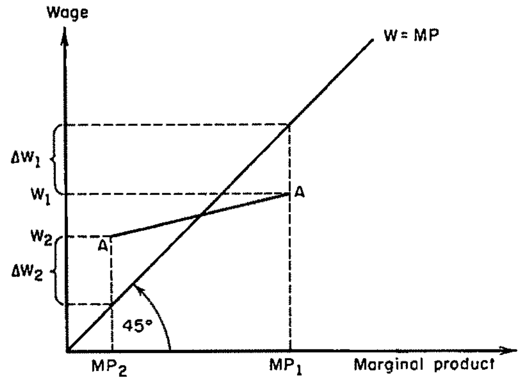


FIGURE 1. AN ATTENUATED PIECE-RATE WAGE SCHEDULE

one worker should associate with another who is less productive depends on whether the value of the status thereby gained by the former exceeds the value of the status lost by the latter. If willingness to pay for improvements in status varies across individuals, I have shown elsewhere (1983, Proposition 2) that wage contracts exist that will cause heterogeneous associations of individuals to form in which status-seeking individuals transfer resources to others who care less about status. These contracts produce Pareto improvements in welfare in comparison with the alternative outcome in which homogeneous groups of workers have wages set equal to marginal products. One's status within one's earnings hierarchy emerges as a good like any other that is traded in the marketplace.

Exact characterization of an equilibrium labor market partition, conditional on even very simple assumed forms of individual utility functions and marginal productivities, is an exceedingly complex task.[10] The discussion here will be limited to an attempt to provide a brief, qualitative description of what a competitive wage structure might look like in the presence of consumption interdependencies. Let us begin by focusing on all individuals with a given marginal product, say, MP_0, and assume that individuals with

[8]An extensive survey of this evidence appears in my 1985 book.

[9]See Proposition 1 in my earlier paper.

[10]For construction of an equilibrium partition based on a simple representation of preferences over status, see my earlier paper, Section III.

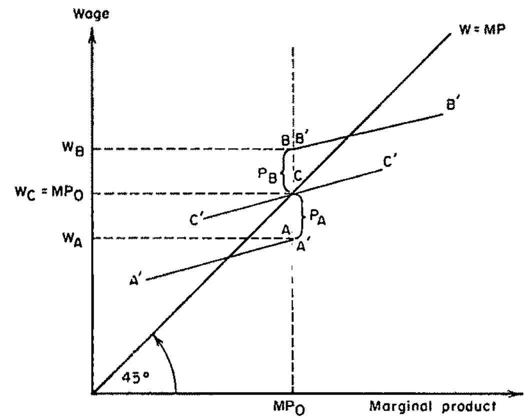


FIGURE 2. LABOR MARKET SUBGROUPS

any given marginal product vary in some positive range in their willingness to pay for improvements in relative standing among their coworkers.

If the total number of workers is sufficiently large, and if the willingness to pay for status improvement is different for each individual with marginal product equal to MP_0, the optimal labor market partition will place each individual in a different labor market subgroup according to how much status he is willing to pay for. If the number of individuals with $MP = MP_0$ is large, the individual in that category who is willing to pay the most for status will, by paying a larger premium than the others, secure the position of top earner in a subgroup of workers whose marginal products are all less than MP_0. This individual's position is denoted as A in Figure 2, which shows that he pays a premium $P_A = MP_0 - W_A$ in return for being the best-paid member of his group of coworkers, the rest of whom are paid according to a schedule that is qualitatively like $A'A'$ in Figure 2.

The individual with marginal product MP_0 who cares least about pay status vis-à-vis coworkers will be able to exploit this aspect of his preferences by earning a wage premium in return for occupying the tail of the pay distribution in a group of workers whose marginal products all exceed MP_0, and who are paid according to a schedule like $B'B'$. This individual is denoted as B in Figure 2, and the premium he receives is $P_B = W_B - MP_0$.

Individuals who are between A and B in terms of their willingness to pay for status improvement will find it in their interest to choose intermediate subgroups in which they either pay a lower premium than A or receive a smaller premium than B in return for locating at less extreme status positions within a labor market subgroup. Some individual of intermediate status preferences will locate at C in Figure 2, in which his status within his subgroup (which is paid according to a schedule like $C'C'$) is neither so high that he must work for less than his marginal product to sustain it, nor so low that it entitles him to a compensation payment to induce him to remain with the group.

The number of market subgroups, the range of productive ability included by each, and the size of the compensating wage differentials are all determined in a complex way by the distributions of individual preferences and abilities. In labor markets involving large numbers of individuals, there will in general be a rich menu of choices within which preferences regarding status vis-à-vis coworkers can be accommodated.

The model of wage determination sketched above rests on the observation (well-supported by evidence)[11] that the interpersonal comparisons that carry greatest weight for a person are those that involve the people with whom he or she associates most closely. This observation, viewed in the context of the preceding discussion, motivates the following simple testable propositions.

PROPOSITION 1: *For cost-minimizing firms that purchase labor services in competitive labor markets, an employee's wage, W, will be related to his or her marginal product, MP, according to a schedule $W = W(MP)$ in which $dW(MP)/dMP$ takes a value significantly smaller than one.*

PROPOSITION 2: *Other things equal, as the tasks performed by a group involve more sustained and intense interaction and contact*

[11]For extensive citations to this literature, see my 1985 book.

between coworkers, the values taken by dW/dMP must be smaller and/or the variation of earnings values included in the group must diminish.

The next section of the paper examines empirical evidence that bears on the validity of these propositions.

II. Earnings vs. Marginal Productivity

The task of this section will be to construct direct estimates of the marginal productivity of specific employees in a variety of enterprises that might reasonably be assumed to purchase labor service under competitive conditions, and then compare these estimates with the amounts these employees are actually paid. The only other study of which I am aware that has attempted to estimate marginal products directly is by Gerald Scully (1974), in which he tried to measure what individual baseball players were worth to their teams. At the time of Scully's study, however, baseball's reserve clause was still in effect, so there was no reason to believe that players would be paid their marginal products. The industries I shall examine below, by contrast, are highly competitive; if the marginal productivity theory of wages, as conventionally formulated, does not perform well in these cases, we have no reason to expect it will perform any better elsewhere.

A. *Pay Schedules for Sales Workers*

As noted, the complex interactions that take place in most team production activities make it exceedingly difficult to isolate and measure the contribution of any one individual employee to his or her firm's net revenues. For a variety of sales activities, however, such interactions are only minimally present, and it is thus possible to achieve relatively precise measures of the contributions of specific employees. Such is the case, for example, for people who sell automobiles and real estate.

What should competitively determined pay schedules look like for such workers as these in a neoclassical economic world in which interdependent preferences are assumed not to exist? To make matters concrete, let us consider the case of a firm that sells some product that may be purchased from its manufacturer at a constant unit price C. Suppose that the firm may attract sales prospects M by the expenditure of resources Y according to the schedule

$$M = M(Y), \tag{1}$$

with $M'(Y) > 0$ and $M''(Y) < 0$. Sales prospects may be thought of as being attracted by expenditures on advertising or through other such activities that enhance the attractiveness of the firm's offering in the eyes of potential buyers. For simplicity's sake, I shall assume that the firm's product sells for a fixed price P, though none of the conclusions that follow would be altered if the model were generalized to say that the firm faced a downward-sloping demand schedule.

Having attracted a given pool of sales prospects, the firm then distributes these prospects in equal shares to each of its N salespersons.[12] These sales prospects are transformed into final sales according to some probabilistic process that differs for each salesperson. In particular, let us assume that the sales, S_i, made by the ith salesperson depend positively on both his ability, a_i, and on the number of prospects assigned to him, M/N, according to

$$S_i = a_i f(M/N) \cdot M/N, \tag{2}$$

with $0 \leq a_i \leq 1$, $0 < f(M/N) < 1$, $f'(M/N)$

[12] It is possible that, during slack times, sales managers may give preferential treatment to their best salespersons in the assignment of sales prospects. If so, then the extent to which the top salesperson's sales exceed others' would overstate the corresponding differences in productivity. But in the sample of firms I studied, at least several managers stressed their practice of equal treatment in their allocations of sales prospects. Among barbers, whose wages I also attempted to study, the standard practice was in fact to give preferential treatment to the *least* successful barbers. That is, when one barber in a shop managed to establish a larger clientele than the others, shop managers would compensate for that fact by assigning walk-ons preferentially to other barbers. The specific details of this practice were unfortunately so ill-defined as to preclude reasonable estimates of the contributions of individual barbers.

<0, $f''(M/N)>0$, and $dS_i/d(M/N)>0$. The expression $a_i f(M/N)$ may be interpreted as the probability that the ith salesperson will transform any given prospect among his M/N sales prospects into a final sale.

The labor market for salespersons is assumed competitive, so that the wage payment to the least able salesperson hired will be equal to the wage W that this individual could have earned in the most attractive alternative activity open to him. Competition implies that salespersons with ability levels higher than this will capture as rents all of the incremental revenues to the firm that result from these higher ability levels.

Given this characterization of demand and cost conditions, the profit-maximizing firm will choose N and Y to maximize

$$(3)\quad \pi=(P-C)\sum_{i=1}^{N} a_i f(M/N)M/N - Y - WN,$$

where the list of K potential salespersons is indexed so that the ability parameters $a_1,\ldots,a_K$ appear in descending order. Let us assume that the specific functions in the maximand are such that interior solutions for N and Y will exist. Call these solutions N^* and Y^*.

For these values N^* and Y^*, the amount by which revenues net of nonlabor costs would decline if the ith salesperson were no longer engaged by the firm is given by

$$(4)\quad MP_i=(P-C)\left[a_i f\left(\frac{M^*}{N^*}\right)\frac{M^*}{N^*} - \sum_{j\neq i}\left(a_j f\left(\frac{M^*}{N^*-1}\right)\left(\frac{M^*}{N^*-1}\right) - a_j f\left(\frac{M^*}{N^*}\right)\frac{M^*}{N^*}\right)\right],$$

where $M^*=M(Y^*)$. The summed expression on the right-hand side of equation (4) represents the increment in sales the remaining N^*-1 salespersons would generate as a result of their each having a larger sales prospect pool to work with in the absence of salesperson i. If we denote this increment as D_i, we may then rewrite equation (4) as

$$(5)\quad MP_i=(P-C)[S_i-D_i].$$

To sustain its operations the firm must recover the costs Y^* it has incurred in generating its pool of sales prospects, $M^*=M(Y^*)$. The firm can remain in business by paying each salesperson MP_i according to equation (5), provided $Y^*\leq(P-C)\Sigma_{i=1}^{N^*}D_i$. I shall assume that entry is free into the market being served by our selling firm and that price will adjust so that $Y^*=(P-C)\Sigma_{i=1}^{N^*}D_i$, resulting in zero profits for the firm. With a wage schedule such as the one in equation (5) in place, workers who could sell at least S_{N^*} units per time period, where $(P-C)(S_{N^*}-D_i)=W$, would find sales work attractive while all others would presumably choose alternative employment at the wage W. Because the wage schedule in equation (5) is constructed under the assumption that there exist no interdependencies in worker preferences, let us refer hereafter to it and other similarly constructed wage schedules as "individualistic" wage schedules.

The foregoing analysis of the sales firm assumes that the only interaction between salespersons that is relevant for purposes of assessing marginal productivity is the one in which the simple presence of each salesperson results in fewer sales prospects being available to the other salespersons in the firm. Now, it is easy to envision a variety of other ways in which the presence of one salesperson might affect the amounts sold by the others. It is possible, for example, that a ruthlessly aggressive salesperson will achieve high sales volumes, but in the process antagonize many of the prospects with whom he comes in contact. The result could be to create an unfavorable reputation for the firm, which in turn would diminish its future flows of new sales prospects. In this case, equation (5) would be an overstatement of the salesperson's marginal product. But it is also easy to envision interactive effects that would cause equation (5) to understate the marginal

products of a firm's most successful salespeople. The presence of skillful salespersons seems, on balance, more likely to enhance, rather than detract from, the firm's reputation, and there is the additional possibility that salespeople of lesser skills will learn to improve their sales techniques by associating with particularly skillful salespersons. Such considerations as these appear to weigh in favor of calling equation (5) an underestimate of the marginal product of the best salespersons, though it will of course be important to bear in mind the possibility that this may not be so when deciding how much confidence to place in the empirical findings I report below.

Given this characterization of what competitive wage schedules should look like in a world without interdependent preferences, let us now examine the wage schedules we observe in practice for activities of the sort at issue.

B. *Automobile Sales Commissions*

Conditions in the automobile dealership industry closely approximate those assumed in the model that led to equation (4). The franchise dealer provides an inventory of demonstrator vehicles, maintains a service facility, and undertakes numerous other expenditures for the purpose of attracting a pool of potential sales prospects. Almost all of these expenditures vary positively with the size of the prospect pool the dealer attracts. For a prospect pool of given size, however, these expenditures may be properly thought of as fixed, irrespective of the number of sales actually made. If P_j represents the retail selling price of a car of type j and C_j represents the wholesale cost of that car to the dealer, let us refer to $G_j = P_j - C_j$ as the dealer's "gross margin" on sales of the jth car type. Assuming each salesperson is given equal access to the pool of potential sales prospects, it follows that when one salesperson consistently sells an average of k cars per month more of type j than does a second salesperson, the first salesperson's contribution to the dealer's net revenues exceeds that of the second salesperson by an amount kG_j per month. If the dealer sells T types of cars, let us use $D_1,\ldots,D_T$ to represent changes in sales volumes for the various car types sold by existing salespersons that occur when an additional salesperson is added to a dealer's staff. We may then express the individualistic commission schedule for automobile salespersons by

$$E_i = -\sum_{j=1}^{T} G_j D_j + \sum_{j=1}^{T} G_j S_{ij}, \tag{6}$$

where E_i denotes earnings per period and S_{ij} denotes sales per period of cars of type j by the ith salesperson.

In order to investigate how automobile salespersons are actually compensated, I collected data on the parameters of the earnings schedules for a random sample of thirteen large automobile dealerships located in upstate New York. If we use R_i to represent the term $\Sigma_{j=1}^{T} G_j S_{ij}$ from the right-hand side of equation (6), then the actual earnings schedules observed for each of these firms was of the form

$$E_i = \alpha + \theta R_i, \tag{7}$$

where the term α was for each case positive if taken to include an imputed value for the "free" use of demonstrator vehicles to which all salespersons in the sample were entitled.

A test of Proposition 1 from the previous section (which says $dW/dMP < 1$) translates in the context of equation (7) as a test of the null hypothesis that $\theta = 1$. The actual values of θ are reproduced for each of the thirteen dealerships in Table 1.

As the entries in column (3) of Table 1 indicate, none of the dealerships in the sample had values of θ larger than .3. The average value for θ was .236, and its standard deviation only .0405. If we assume the values of θ to have been drawn from a normal population, then the relevant t-statistic is given by -18.87, which enables us to reject the null hypothesis at conventional significance levels.

Perhaps it is the case that the observed commission schedules for automobile salespersons are so much flatter than the individualistic schedules because salespeople have

TABLE 1—SLOPES OF EARNINGS FUNCTIONS AND MEASURES OF EARNINGS VARIABILITY FOR THIRTEEN AUTO DEALERSHIPS

Dealership (1)	Number of Salesmen (2)	θ Slope of the Earnings Function (3)	$\frac{\mathrm{Min}_t(E^*_{i_t})}{\bar{E}}$ (4)
1	15	.25	1.2
2	10	.25	1.5
3	10	.20	1.5
4	14	.20	0.9
5	8	.30	1.0
6	10	.30	1.2
7	10	.25	1.2
8	11	.25	1.6
9	9	.20	1.5
10	10	.25	[a]
11	8	.20	[a]
12	14	.25	[a]
13	17	.165	[a]

[a]Not available.

highly variable sales records and prefer gently sloped earnings schedules as a means of insulating themselves from exposure to risk. Or, perhaps the average sales volume for most salespeople lies very close to their respective group's average, so that while individual earnings figures might differ from marginal products in any given time period, an individual's average earnings value through time nonetheless lies very close to the average value of his marginal contribution to his dealer's net revenues. Several considerations appear to rule out these alternative interpretations of the observed earnings schedules. Every dealer in the sample reported that there was a high degree of consistency in the rankings of annual sales figures for their long-term sales employees. Several firms reported that their best salespeople typically had at least 70 percent higher sales volumes than those of the average salesperson. These observations appear to rule out the possibility that there are no identifiable long-term differences in the marginal products of different salespeople.

To explore the possibility that risk aversion might explain the flatness of the observed earnings schedules, I first asked the sales managers of the dealers to identify their most consistently productive salesperson, and then to estimate the ratio of that salesperson's lowest annual earnings figures (in constant dollars) to the average earnings figure for all salespersons during a typical year (also in constant dollars). These ratios are given in column (4) of Table 1. In every case but one, this ratio exceeds one, and the one for which it was less was only slightly so, and represented last year's earnings figures, which the dealer described as very severely depressed in relation to long-run averages. (This dealer had the misfortune of owning an Oldsmobile franchise in the midst of the latest escalation in gasoline prices.) To illustrate what these responses seem to say about the risk-aversion explanation, let us plot the individualistic earnings schedule from equation (4) and the observed earnings schedule for dealer 2 in Figure 3, where $\mathrm{Min}_t E^*_{i_t}$ and $\mathrm{Min}_t R^*_{i_t}$ denote the lowest values taken in any year by dealer 2's best salesperson's earnings and gross margin values, respectively.

The zero-profit constraint requires the average commission value to lie on both the individualistic and observed earnings schedules. These schedules intersect at the point $(\bar{E}, \bar{R})$ in Figure 3, which lies to the left of the minimum value of the best salesperson's earnings. Can it make any sense to say that risk aversion makes the most productive salesperson choose the observed compensation schedule, the minimum point on which for him is $\mathrm{Min}_t E^*_{i_t}$, in preference to the individualistic schedule? That would be to say, preposterously, that that salesperson prefers one income stream to another that, while more variable, is also element-for-element larger than the first stream. A better explanation than that seems needed for why the best salespeople have not been bid away from existing firms by new firms whose compensation schedules have θ closer to one than in the schedules of existing firms.

C. *Real Estate Sales Commissions*

Real estate sales commissions are slightly more complicated than automobile sales commissions in that most real estate firms act as participants in local multiple listing services. A typical house for sale on the

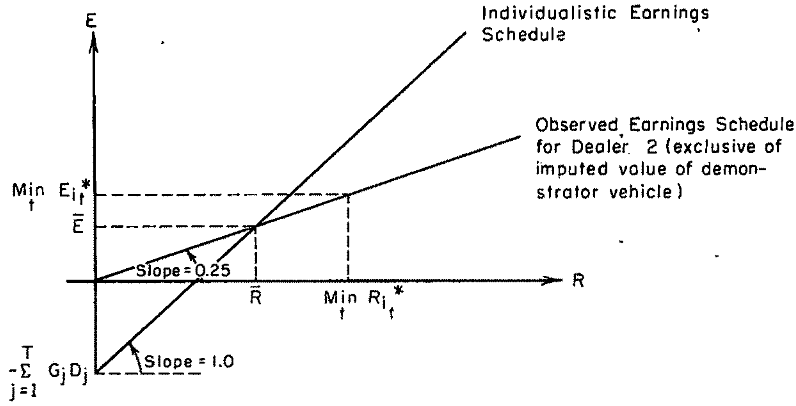


FIGURE 3. THE LOWEST EARNINGS OF THE BEST SALESPERSON

market will be listed with a particular agency and at the same time listed with the multiple listing service (MLS). For a house to appear in the MLS means that all member agencies receive information on the price and other relevant market characteristics of the house, and an implicit invitation to show the house to their clients. When a client of the listing firm purchases a house that also appears in the MLS, that firm receives all of the resultant sales commission, which is some share μ of the contract price of the house. Typical values for μ lie between 6 and 7 percent for residential properties. When the house is purchased by a client of some other member agency of the MLS (a "cobrokered" sale), that agency receives a share λ of the total sales commission, with the remaining $(1-\lambda)$ share going to the original listing agency. Typical sharing schemes give 60 percent of the total sales commission to the selling agency and 40 percent to the listing agency.

In a large market in which many real estate firms participate in the MLS, the modal sales transaction involves a client of one firm purchasing a house listed by some other firm. The firm that makes the sale typically will not have expended significant resources advertising the house in question, and most of the costs that are marginal to the selling agency for such transactions are ones borne by the selling agent herself (primarily for the running and maintenance costs of the car she uses to transport clients who wish to inspect houses listed by other firms).

For sales by an agent in one firm of houses listed by other firms, then, the conditions that give rise to a wage schedule like the one in equation (4) all appear to be met. If we use V_i to denote the sum of the selling prices for all of the houses sold by agent i in such transactions, then the relationship between V_i and the sales commission E_i paid to the agent who generates the revenues the firm garners from the sales in question should be

$$E_i = -\lambda\mu D_i + \lambda\mu V_i, \tag{8}$$

where D_i, as before, represents sales the remaining agents would have made had they had access to agent i's sales prospects.

When an agency acquires an additional listing, it incurs a number of expenses that it would not otherwise have incurred, principally advertising outlays for the purpose of calling the house's availability to the attention of the market. For a given number of houses listed, however, the agency's total costs will be approximately the same, irrespective of the number of its listings actually sold by its own sales force. Thus the case of houses sold by the listing firm may be treated in much the same way as cobrokered sales. For brevity's sake, however, let us focus on the case of the cobrokered sale, which in any event is, as noted, the most common sales transaction.

How do agencies actually compensate their sales agents for making cobrokered sales? In an attempt to answer this question, I col-

TABLE 2—COMMISSION RATES PAID BY FOUR REAL ESTATE FIRMS

	Gross Commission Rate μ	Cobroker's Commission Rate λ	Listing Agent's Commission Rate ϕ	Selling Agent's Commission Rate θ
Firm 1	.06	.4	.50	.50
Firm 2	.06	.4	.50 if $E < \$30{,}000$ .55 if $\$30{,}000 < E \leq \$41{,}111$ .575 if $\$41{,}111 < E$	same as ϕ
Firm 3	[a]	[a]	.50	.50 if $E < \$26{,}000$ .60 if $\$26{,}000 < E \leq \$38{,}000$ .70 if $\$38{,}500 < E$
Firm 4	[a]	[a]	.50	.50 if $E < \$12{,}000$ .55 if $\$12{,}000 < E \leq \$18{,}000$ .60 if $\$18{,}000 < E$

[a]Not available.

lected data from four large real estate agencies located in Ithaca, New York. The Ithaca real estate market is not a large market by any means, but there are, nonetheless, 32 member agencies in the city's multiple listing service, and the Ithaca Board of Realtors reports that the majority of transactions consummated by these member firms are cobrokered sales.

Firms in this sample pay two different types of commissions to agents: listing commissions and sales commissions. All of the houses listed with a given firm are assigned to individual agents within that firm. Some of these listings originate when agents answer the agency's telephones and others result from the personal contact networks of the agents. When a house is sold, its listing agent automatically receives a share ϕ of her firm's listing commission, no matter who sells the house. If the agent herself sells her own listing, she receives both her listing commission and a sales commission, which is a share θ of her firm's selling commission. If she sells another firm's listing, she receives the same share θ of her firm's selling commission, which, as noted earlier, is a $\lambda\mu$ share of the selling price of the house.

Values for μ, λ, and ϕ, and θ for residential sales for the four firms in the sample are listed in Table 2. For some of the firms, the values of ϕ and θ depend on the total volume of an agent's sales commissions, as indicated in the table.

Of particular interest for present purposes are the values of θ in Table 2, the agent's sales commission rate for cobrokered sales. The claim that wages for agents are determined competitively and that agents' preferences are not interdependent corresponds in the current framework to the claim that θ is equal to one. None of the values of θ observed for any of the four firms in the sample is even close to one. Other available evidence suggests that the values of θ observed in the sample are typical of those for other real estate firms.[13] Even if our information were limited to the four firms at hand, and we took the highest value of θ observed for each firm, we could still reject the null hypothesis that θ equals one at the 1 percent significance level.

For reasons similar to those given in the case of automobile salespeople, employee risk aversion does not seem a plausible explanation for the flatness of the agent compensation schedule in our sample of firms. While firm owners would not release detailed individual earnings records to me, owners did report in each case that their best salespeople had earnings levels that were several multiples of those of their least productive salespeople and that these earnings differentials were highly consistent over time.[14]

[13]See, for example, National Association of Realtors (1982).

[14]Unfortunately, for neither the automobile nor real estate firms I studied was I able to obtain the detailed

D. *Pay Schedules for Professors in Research Universities*

1. *A Thought Experiment*. If the measurement of employee productivity in sales industries may be said to lie near one end of the difficulty spectrum, then the same task for professors in research universities must surely lie close to the other. The process whereby individuals and groups of professors labor to produce "knowledge" through their collegial interactions, research, teaching, and administrative duties appears so bewilderingly complex as to defy any reasoned effort to define, much less measure, the specific contribution made by any individual employee.

Despite the obvious difficulties involved in assigning value to the contributions of specific employees, however, it is important to bear in mind that many thousands of such judgments are made every year. Major research universities conduct their operations in what may be fairly described as a harshly competitive environment. In such an environment, let us ask what sorts of personnel judgments a decision maker would have to be able to make in order to protect his or her group's position in the academic hierarchy. Let us take as a working hypothesis that there is very substantial overlap in the range of productive contributions made by scholars in departments that are ranked close together in whatever commonly accepted hierarchy may exist in a given field. Let us also suppose that people are willing to move in response to a better salary offer made by an institution of comparable rank to their own, or are at least willing to affect such a posture in order to induce their respective deans to match such salary offers.

Given these assumptions, a particular department's observed stability in the academic hierarchy implies that neither its own decision makers, nor those of its closely ranked competitor departments, are able to identify members of other departments who are underpaid in relation to the true values of their productive contributions. If that were not the case, some department could have improved its position by having hired these underpaid members of a competitor's department in lieu of whatever alternative hiring or promotion moves they last took.

If the reader is a participant in personnel decisions in an academic institution, he or she is then well positioned to undertake a thought experiment to test the validity of the foregoing claim that observed stability in a hierarchy of departments implies that professors in those departments are paid their marginal products, at least within the limitations that the relevant decision makers face in making such judgments. To carry out this thought experiment, let the reader first identify those two of his colleagues in a given age group whose contributions to his department's mission he considers more valuable than those of any other two colleagues in the same age group. Let the reader then identify those three colleagues in this same age group whose combined efforts make, in his view, the least valuable contribution to his department's mission. If the age and salary structures of the reader's department resemble those reproduced in Table 3 for members of the Department of Economics at the University of Michigan, it will be the case that the three professors in the second group will together earn more than the two in the first group. (In most departments, this statement is true for virtually any groups of three and two professors in the same age group, and is meant to imply no judgment about the quality of any particular economist at the University of Michigan. If the statement is not true for the reader's department, then the conditions for carrying out the thought experiment do not apply.) The concluding step in the thought experiment is for the reader to decide whether the absence of the first group of two professors would more seriously detract from his department's mission as he

individual earnings data that would be needed to determine whether a stratification process of the sort described in Section I takes place in these industries. Among the people I interviewed, however, there was general agreement that a small number of elite firms exists in each industry, to which many of the best salespersons eventually gravitate. In the real estate industry, for example, the payoff to being associated with such a firm is that its listings tend to come disproportionately from upscale neighborhoods.

p14421

TABLE 3—EXPERIENCE AND SALARY STRUCTURE OF UNIVERSITY OF MICHIGAN ECONOMISTS, 1983–84

Faculty Member	Experience[a]	Salary	Faculty Member	Experience[a]	Salary
1	43	$63,000	17	18	$45,000
2	32	54,300	18	17	50,700
3	32	51,000	19	17	37,500
4	30	39,000	20	16	61,000
5	26	52,000	21	16	48,100
6	25	55,000	22	16	30,000
7	23	41,200	23	15	51,500
8	22	47,700	24	13	40,600
9	22	44,500	25	12	51,300
10	21	43,000	26	12	50,300
11	20	46,800	27	10	62,400
12	20	42,400	28	10	39,300
13	19	56,500	29	9	43,200
14	19	55,000	30	7	40,400
15	19	53,000	31	6	37,700
16	18	55,000	32	5	27,700

[a]Years since receiving Ph.D.

sees it than would the absence of the second group of three professors. If professors are paid in proportion to their marginal products, the answer to this question should be no. An unequivocal yes to this question, on the other hand, may be interpreted as further support for the claim that the distribution of employee compensation shows less dispersion than does the distribution of their marginal products.

It is possible that the vagaries of the tenure system may cloud the results of the foregoing thought experiment in some instances. Some faculty members, for example, may enter a period of quasi retirement upon being granted tenure, yet be protected by the tenure system from having their wages adjusted fully to reflect the new, diminished levels of their productive contributions. The salaries of such faculty would of course be greater than their marginal products, but that tells us nothing about the issues under investigation here. To abstract from such essentially extraneous information, the foregoing thought experiment should be confined to those faculty to whom the granting of tenure is not viewed in hindsight as having been a mistake.

2. *The Value of Grantsmanship*. In an effort to generate evidence of a less subjective character on the question of whether salaries track productive contributions in academia, I gathered data on the research grants obtained by a sample of eighteen biochemists and organic chemists, all full professors at Cornell University.

When a Cornell professor obtains a research grant from an outside source, the university's grant contract stipulates that the donor shall pay the direct costs itemized in the accepted proposal's budget, plus an additional contribution to the indirect costs of operating the university's infrastructure. The standard value of this indirect cost contribution is 49 percent of all budgeted direct costs. Contributions much smaller than that figure are routinely negotiated with individual donors, however, and many donors make grants to Cornell that make no indirect cost contribution at all. The university is reluctant to disclose the identities of donors whose grants make no contributions to indirect costs, apparently out of concern that to do so might compromise its ability to garner indirect cost contributions from other sources. My discussions with the university's Office of Sponsored Programs made it clear, however, that the university would not hesitate to accept an additional grant that made no contribution to indirect costs if the alternative were to miss out on that grant entirely, and if its acceptance would not alter

TABLE 4—INDIRECT COST INCOME GENERATED BY CORNELL ORGANIC CHEMISTS AND BIOCHEMISTS

Group	1979–80 Indirect Cost Income per Professor
Top 3	$245,000
Median 4	92,994
Bottom 3	8,575

TABLE 5—1979–80 RESEARCH SUPPORT LEVELS VS. THREE-YEAR AVERAGES

Group	Individuals Ranked by 1979–80 Research Support	1979–80 Support / 3-Year Mean Support
Top	1	.883
	2	1.222
	3	.925
Median	8	.890
	9	1.211
	10	1.163
	11	.879
Bottom	16	.622
	17	.565
	18	[a]

[a]No support during previous three years.

the indirect cost contributions it receives from other donors.

My working hypothesis in the discussion that follows will be that the university's position on this issue is well-founded; that is, I shall assume that even those research grants that make no contributions to indirect costs contribute in other ways to the enhancement of the university's academic mission. Given this assumption, it then follows that in the process of bringing a research grant to Cornell, a professor's contribution to the university's mission (however that mission might be defined) must be at least as great as the indirect cost contribution that accompanies that grant.

For expositional convenience let us focus on the top three, the median four, and the bottom three grant recipients in the sample of eighteen professors for the academic year 1979–80. The per capita indirect cost contributions that accompanied the 1979–80 grants brought in by these three groups of professors are as listed in Table 4.

With the possible exception of the indirect cost figures for the bottom three grant-getters, the figures in Table 4 are in substantial measure indicative of consistent differences in the contributions to indirect costs made by the grants received by these professors. For each of the professors listed in each of the three groups in Table 4, the ratio of total grant support for 1979–80 to total grant support for the preceding three-year period is listed in Table 5. As indicated by the entries, the 1979–80 contributions to indirect costs for professors in the top and median groups were all very close in value to the three-year average figures. Two of the three professors with the least funding in 1979–80 had significantly higher average funding levels for the previous three years, while the third had received no outside research support at all during that same period. Even for the bottom group, then, there is no clear indication that the 1979–80 research figures are dramatically out of line with that group's long-run research funding levels.

In further confirmation of the impression given by the entries in Tables 4 and 5, I was informed by the office of the associate dean of Cornell's College of Arts and Sciences that the entries in Table 4 are approximately representative of the long-run funding level differences between the three groups in question. I shall assume in what follows that the figures in Table 4 do indeed represent long-run differences in the indirect cost contributions generated by these groups of professors from outside sources. But the general conclusions that follow from this assumption would not be altered even if the top group's figure were half as large, and the bottom group's twice as large, as the corresponding figures in Table 4.

Unlike the University of Michigan, which is required by state law to make its faculty salaries public, Cornell keeps its own faculty salaries a carefully guarded secret. The associate dean's office was willing to reveal, however, that the salaries of organic chemists and biochemists at Cornell vary within a range that is not more than 10 percent different at either end from the range of faculty salaries in chemistry at the University of Michigan. (I was told, interestingly, that

TABLE 6—SALARY DIFFERENCES BETWEEN GROUPS

Ranking[a]	Cornell Salary Index	Estimated Dollar Salary in 1980–81
Top	1.00	58,025
Median	.92	53,383
Bottom	.70	40,618

[a]By amount of research support.

TABLE 7—SALARY DIFFERENCES VS. DIFFERENCES IN INDIRECT COST CONTRIBUTIONS

Group Comparison	Average Difference in Salary	Average Difference in Indirect Cost Contribution
Top vs. median	$4,642	$152,006
Median vs. bottom	12,765	84,419

salaries in biochemistry are much higher at Michigan than at Cornell because "Michigan's medical school is located right in Ann Arbor." The close interactions that ordinarily take place between highly paid medical school professors and biochemists do not develop at Cornell, whose medical school happens to be located in New York City.) Given the relationship between Cornell and Michigan salaries, the average salary of the three best-funded professors in the Cornell sample cannot be larger than $58,025, the figure that is 10 percent higher than the salary of the most highly paid full professor in the chemistry department at the University of Michigan. The associate dean's office was also willing to reveal the relative values of the average salary levels for the three groups of professors. Combining these two pieces of information produces the estimates shown in Table 6 of the average salary levels for the three groups of Cornell professors. For our purposes, the salary estimates in Table 6 may be called conservative, in the sense that they overstate the true differences in salary levels between the three groups in question.

Using the entries from Tables 4 and 6, let us next compute the differences between groups in average salary levels and average contributions to indirect costs. These differences appear in Table 7.

Viewed from the perspective of the university, which has ultimate authority over faculty salary policy, the differences in indirect cost contributions shown in Table 7 represent one important dimension along which marginal products differ between the three groups of professors at issue. These marginal product differences are dramatically larger than the corresponding figures for salary differences, which as noted, are themselves overestimates of the true salary differences between the three groups. It is difficult to argue that differences along other dimensions of productivity could possibly be large enough to offset the observed differences in indirect cost contributions. As in other fields, it is presumably also true in the sciences that a scholar's ability to obtain research grants is positively related to the quality and quantity of his or her past research efforts. Also to be considered is the fact that the direct cost portion of most grant budgets contains many expenditures—for the purchase of capital equipment and support for graduate students, to name two—that make additional contributions to a department's academic mission. Finally, while faculty with extensive outside research support are often observed to teach fewer courses than other faculty, this fact is largely offset by funds allocated within the direct cost portion of grant budgets for the specific purpose of purchasing released teaching time.

These are all points that could be investigated explicitly if the names of the individuals in the specific groups in Table 7 were known to us. Since they are not, let us simply assume, for the sake of advancing the discussion, that productivity along other dimensions than grant-getting is inversely related to grant-getting for our sample of professors. The most extreme assumption we can make is that the top group contributes nothing to the university except the grants it receives, that members of the bottom group earn their salaries exclusively on the basis of contributions in areas unrelated to grant-getting, and that members of the median group make some contribution to the university other than the grants they receive. Given this extreme set of assumptions, the marginal productivity theory of wages would presumably

say that the average value of the bottom group's contributions along nongrant-related dimensions is equal to its average salary, which we have seen is a number less than \$40,618/year. This in turn implies that the differences in marginal products between professors in the top and bottom groups must be at least as large as \$195,807 (which is the difference between average indirect cost contributions for the two groups, less the average salary of the bottom group). The difference in average salary levels for the top and bottom groups, on the other hand, is no more than \$17,417, or less than 9 percent of the former figure.

Perhaps universities are reluctant to allow salaries to reflect indirect cost contributions generated by faculty for fear the flow of grants might suddenly be discontinued, leaving them in the always-difficult position of having to negotiate reductions in existing salary levels. The professors in the top group, however, would presumably be willing to stipulate in advance that their salaries be linked explicitly to their grant totals, if the alternative were to be forced to settle for their current salary levels. More to the point, a competing university that offered a 50-year-old top grant-getter a mere \$10,000 premium over his or her existing salary would only need to see one year of successful grant-getting to have its gamble pay off. Even the most risk-averse university would be unlikely to walk away from such an apparently favorable gamble as that.[15]

Quite possibly the granting agencies themselves might object if universities were to link professors' salaries by formula to the amount of indirect cost contributions they managed to generate. The possible existence of such a constraint seems plausible indeed, since the express purpose of indirect cost contributions is for the granting agency to pay its fair share of the university's overhead expenses. Yet the fact that such a constraint may exist in principle does not imply that it is binding at currently observed salary levels. To be sure, it would be difficult to argue that any granting agency would have objected or altered its grant awards in any way had some competing university managed to hire one of Cornell's top grant-getters on the strength of a \$70,000 annual salary offer. That such offers were not forthcoming may be taken as a signal that some factor other than external constraints from the granting agencies must explain why the indirect cost contributions generated by the professors in the sample are so weakly reflected in their salary differences.

One might reasonably interject at this point that if all indirect cost contributions were captured as rents in professors' salaries, the infrastructure upon which a university's reputation so strongly depends would deteriorate, which would in turn seriously diminish the capacity of its professors to attract financial support from outside sources. Though this observation may be correct, it poses no challenge to the interpretation of the internal wage structure offered here. With respect to the receipt of outside grants of financial assistance, university infrastructure expenses are properly viewed as fixed costs: the costs of libraries, philosophy departments, snow removal equipment, and administrators' salaries are not materially different when some professor receives an additional research grant from what they would have been had she or he not received that grant. Any university that managed to hire away one of Cornell's top grant-getters for an annual salary of \$70,000 would capture at least \$175,000 in *additional* annual funding to help defray the costs of maintaining its own infrastructure. In the absence of a grand conspiracy theory of university wage determination, the fact that serious adjustments might be required if *all* universities paid faculty salaries that fully reflected the indirect cost contributions they generate cannot be offered as an explanation of why it is not strongly in the interests of any one institution to bid vigorously for the services of proven

[15]A related risk is that a university might have sufficiently many grant-getters that its total indirect cost collections would exceed an amount deemed reasonable by government funding agencies. Such a university would then no longer stand to gain in the same way that others would by the hiring of a prolific grant-getter. But Cornell was not, and does not expect to be, confronted with the problem of too many grants. The problem for even the most well-funded universities has rather been their failure to attract sufficient indirect cost support for the research programs they carry out.

grant-getters. Yet such bidding, to the extent that it has taken place, has fallen far short of the result that would have been predicted by the traditional individualistic theory of wage determination.

I argued at the outset that an implicit market for status in each firm's internal earnings hierarchy causes the distribution of earnings to be compressed relative to the distribution of marginal products. But it strains credulity to assert that all major grant-getters at a university could feel so strongly about occupying a high-status position that they are willing to sacrifice hundreds of thousands of dollars per year for the privilege of doing so. Why don't those proven grant-getters who care little about status leave their current posts in favor of employment in an alternative setting in which their status is lower but their income is closer to the value of what they contribute? One coherent answer to this question is that production complementarities may stand in the way of organizing an economically viable firm in which top grant-getters do *not* occupy very high positions in the income hierarchy. If we assume that for a university to achieve genuine distinction it needs strong performance not only in the sciences, but in arts and humanities as well, and if we assume that grant-getting possibilities outside the sciences are relatively limited, it then follows that a top grant-getter simply cannot be paid his marginal product without earning vastly more than most other university faculty members. Those at the very top of a population's marginal product distribution may be forced, then, by the lack of a viable stand-alone option, into spending much more for status than they would if a broader menu of choices were available.

E. *Variations in the Price of Status*

Proposition 2 from Section I holds that, other things equal, as coworkers work in closer contact with one another on the job, we should expect to see the slope of the earnings schedule become flatter, or to see the range of earnings within the group become more limited. Reduced earnings variability is, of course, one consequence of having a flatter earnings schedule, but can be achieved also by limiting the range of productive abilities included within the relevant group of coworkers.

I offer no scientific evidence here that automobile salespersons have closer ties to their coworkers than do real estate salespersons, other than to observe that the former spend most of their time working in close physical proximity of one another, while many of the latter work primarily out of their own homes; and most spend the bulk of their working hours visiting houses with prospective buyers. Both the psychologist's models of attention and the sociologist's description of the reference group stress the roles of exposure and proximity as determinants of what we focus most closely upon.[16]

On the basis of these considerations, then, we expect the implicit price of status to be higher among people who sell automobiles than among those who sell houses. As we have already seen from the entries in Tables 1 and 2, the slopes of the observed earnings schedules are indeed substantially smaller for autosalespersons than for real estate salespersons. As concerns the range of earnings variations within groups, the information I was able to gather is relatively more fragmentary, but what there is does accord with the notion that the difficulties created by a given degree of earnings dispersion in a work group increase with the degree of contact that occurs between the group's members. For three of the four real estate firms for which the relevant information is available, the maximum annual earnings averaged 5.7 times that of the minimum annual earnings for full-time salespeople. Among auto dealerships, the corresponding ratio was only 3.1 to 1.

In a similar vein, it is plausible to argue that research professors in the sciences interact more intensively with one another than do coworkers in either of the two sales oc-

[16] In characterizing the information people rely on in making judgments, psychologists stress the importance of knowledge that is "salient," "vivid," and "available." Information acquired in close proximity to actual events is rich along all of these dimensions. See Richard Nisbett and Lee Ross (1980, ch. 3); S. E. Taylor and S. T. Fiske (1978); and Amos Tversky and David Kahneman (1973). See also Robert Merton and Alice Kitt (1950), Leon Festinger (1954), and Robin Williams (1975).

cupations discussed above. In addition to working in close physical proximity to one another, it is common for professors to engage in collaborative research projects, especially in the sciences. They frequently coauthor books and articles, are often members of the same administrative committees, and frequently cooperate in supervising the same students. Moreover, job turnover is much less frequent among full professors in universities than among automobile or real estate salespersons (the median length of tenure to date for real estate salespersons in my sample was less than two years), which means that professors come in contact with one another as colleagues over a longer period of time than do people in the two sales occupations. For the research professors in the sample, we saw that the top earners received less than 50 percent more than the bottom earners. We saw further that salaries of professors might go up by as much as nine cents for every extra dollar they contribute to the university's net revenues, but probably go up by much less than that. These observations also support the notion that the price of status is highest in groups whose members interact most intensively.

In addition to the relationships discussed above, we have available one other piece of information that sheds light on the question of how the extent of contact between coworkers affects earnings variability within firms. Jacob Mincer (1981) has found that the average length of job tenure is substantially longer for union workers than for nonunion workers,[17] a finding I shall interpret to mean that the average degree of contact between coworkers is higher in union than in nonunion firms. A large literature also supports the finding that rates of pay vary much less with respect to standard proxies for productivity in union than in nonunion firms,[18] just as predicted by Proposition 2.

Given only the foregoing observations to work with, it hardly needs saying that findings on how the price of status depends on the closeness of contact between coworkers are little more than suggestive at this stage. Still, if the reader accepts my closeness-of-contact rankings for the various groups of coworkers for which information on pay variability is available, these observations nevertheless have a certain measure of statistical force. Letting X_1, X_2, X_3, X_4, and X_5 represent the slopes of the earnings schedules for real estate salespeople, auto salespeople, research scientists, nonunion workers, and union workers, respectively, Proposition 2 then predicts the occurrence of the compound event $\{(X_1 > X_2 > X_3) \text{ and } (X_4 > X_5)\}$. Working within a nonparametric framework, as seems natural here, let us state as our null hypothesis that $X_1, \dots, X_5$ are independently, identically distributed. The probability of the compound event predicted by Proposition 2 having occurred by chance is then only .083, which is within the conventional limits used for rejecting hypotheses in economics.

III. Additional Considerations

A. *Limits on Piece-Rate Earnings*

There is evidence that the pattern of earnings compression observed in the cases examined here seems to be widespread. In a survey of sales commission plans, whose results are summarized in Table 8, the National Industrial Conference Board (1970) reports that 54 out of the 100 plans studied actually imposed ceilings above which a salesperson's earnings could not rise, *irrespective of his sales output*. Measurement difficulties and employee risk aversion can hardly be expected to explain the existence of this type of limitation on the amount employees are permitted to earn.

A large literature exists describing the widespread practice, in both union and nonunion firms, in which workers impose strong sanctions against those of their number who exceed clearly stated norms for maximum piece-rate production.[19] This practice is often explained by writers in the industrial pay

[17]Mincer finds that quit rates in the union sector are about one-half as large as in the nonunion sector for young men, and about one-third as large for men over 30.

[18]See, for example, Mincer, Table 12. See also Richard Freeman (1980).

[19]See, for example, F. J. Roethlisberger and W. J. Dickson (1972), R. Marriott (1957), Wilfred Brown (1962), William Wolf (1957), and Garth Mangum (1964).

TABLE 8—INCENTIVE COMPENSATION CEILINGS BY TYPE OF SELLING ACTIVITY

	Total Plans	Sales Engineers	Type of Selling Activity: Direct Contact	Type of Selling Activity: Promotional Selling
Have no Ceiling	46	17	20	9
Have Ceiling	54	16	20	18
Total	100	33	40	27

literature as being motivated by workers' fears that management will reduce piece rates if workers produce too much.[20] Yet contracts could easily be written that would allay such fears, and the literature reports numerous instances in which management is fully aware of the quotas workers impose on one another.[21] We even know of cases in which limitations on piece-rate earnings come at the initiative of management itself. Robert McKersie describes the case of a General Electric plant that abandoned an incentive pay experiment, despite its strong effect on productivity, because it resulted in some production workers earning more than their supervisors. He reports another instance in which a large manufacturing firm refused, for similar reasons, a union request that it increase the speed of its production assembly line. The marginal productivity theory of wages, as conventionally presented, has great difficulty accounting for such observations as these. These same observations, however, are fully in accord with the model of wage determination discussed in Section I above, which itself lies comfortably beneath the cover of the standard neoclassical umbrella.

[20]See, for example, William Whyte (1955).

[21]See, for example, Mangum, Robert McKersie (1967), Donald Roy (1972), and Wolf. Roy, a sociologist who spent eleven months as drill press operator in a steel fabricating plant, reported that his coworkers wasted more than one-third of their time, often in full view of their foremen, in order to avoid exceeding their self-imposed production quotas. Mangum reports the case of a group of manufacturing employees who routinely played cards once they reached their daily quotas, prompting several of their wives to complain to management about their gambling losses. Yet the process continued. Similar instances abound in the industrial pay literature.

B. *Income and the Demand for Status*

If status is like most other goods, people will demand more of it as their incomes rise. Yet the total proportion of high-ranked positions available to the members of any population is by definition fixed. Thus we may expect that the implicit price one must pay in order to occupy a high-ranked position in an earnings hierarchy (or the premium one receives for occupying a low-ranked position) will rise over time as incomes grow. And a steady flattening of the slopes of incentive pay schedules in the United States has indeed taken place during this century.[22]

The question of how the slopes of earnings functions compare, at a given moment in time, across firms with different average productivity levels raises an important question about the extent to which workers are free to choose their coworkers. Where people within firms work independently of one another, as in my sales examples, equilibrium requires that the implicit price of a high-ranked position be the same, irrespective of the firm's average productivity level. If that were not the case, someone would be paying more than necessary for the privilege of occupying a high-ranked position. If, for example, a low-ranked member in a highly productive firm were receiving a premium larger than the premium received by a low-ranked member in a second, less-productive firm, then the latter worker could offer to occupy a low-ranked position in the first firm for a smaller premium than its current occupant receives. The prices of positions of a given rank would thus adjust until they were equal across firms of different average productivity levels.

In most actual firms, however, there is a great deal of interaction between coworkers, and it will not ordinarily be feasible for an employee from an unskilled group to assume a low-ranked position in a group of highly skilled workers. It is hardly likely, in any event, that the presence of an unskilled worker in a group of highly skilled workers would be perceived by the latter as enhancing their status in any meaningful sense. To have high status in a firm means to occupy a

[22]See, for example, McKersie.

high-ranked position among people who might reasonably be deemed to be one's peers. (To suppose otherwise would be to suppose that one's demand for status could be well satisfied by hiring a derelict to sit at a desk in the corner of one's office.)

If both the means and the efficacy of forming disparately skilled work groups are thus severely limited, then it need not be the case at all that the implicit price of high-ranked positions be the same in all groups. On the contrary, if demands for status rise with income, as assumed, we then expect that the slopes of the earnings schedules for groups with high average productivity levels to be flatter than those for groups with low average productivity levels. In practice, it does in fact appear that the closest we ever come to seeing pure piece-rate pay schemes is in the wage schedules of impoverished immigrant sweat shop workers and farm laborers. By contrast, piece-rate schemes of any sort are almost unheard of in the upper reaches of the labor market. Granted, the difficulty in measuring an individual's contribution to his firm's output is often greater among highly skilled groups than among less-skilled groups. But as we saw in the case of research scientists in Section II, reasonable lower bounds can often be assigned to individual differences in productivity even in complex team production processes. Yet we still do not witness piece-rate schemes for such groups.

C. *The Organization of the Firm*

The model of wage determination summarized in Section I also has implications for how enterprises are organized to perform certain tasks. Consider, for example, the conditions under which oil exploration geologists perform their services. In the task of exploring for oil, an unusually talented geologist may literally be worth several times his weight in gold, and is certainly a much more valuable resource in the process of producing oil than is the chief executive officer of the large multinational oil company that will ultimately extract the oil from the earth. If physical production relationships were all that mattered in dictating the organization of firms, it would appear that the most talented exploration geologists would be best utilized as integrated members of the technical staffs of these large oil companies. In fact, however, the most talented finders of oil often perform their services as independent contractors, an observation that might be explained by the understandable difficulty an oil company would experience if it paid a technical staff member a salary many times those received by its most-senior management personnel.

In the same vein, we observe that many firms apparently prefer to maintain ongoing relationships with costly consulting organizations rather than hire the consultants directly to perform the same services within the firm. Numerous plausible reasons can be offered for why firms might thus continue to incur heavy overhead billings that seem avoidable. But the recognition that status in internal earnings hierarchies has high value suggests the possibility that one of the firm's motives may simply be to isolate its own internal wage structure from costly ripple effects that would be stimulated by internal placement of these highly paid consultants.

Similarly, we may note that the internal requirements of a large hierarchical wage structure often appear to preclude certain useful incentive devices from corporate pay schemes. This in turn may help account for why junior corporate officers are often able to form small new firms that compete successfully against their larger former employers in environments in which economies of scale might otherwise be expected to prevent such an outcome. Where roadblocks stand in the way of an able frog's becoming the biggest in his current pond, it may then pay him, and the rest of society as well, to start a new pond.

D. *The Leveraged Buy Out*

Just as it may often pay for the most productive contributors to break away from a given earnings hierarchy, so also it appears occasionally to be in the best interests of a given hierarchy to shed its least productive members. Consider, for example, the case of a conglomerate firm, one of whose divisions has fallen upon hard times. Because of a change in consumer tastes, its product no longer is in as high demand as before. If the

division were a separate firm, the standard response would be to lower the price of its product, together with the effective wage it pays its workers (if not by an actual nominal wage cut, then by a tightening of work rules).

But when the losing operation is an integral part of a conglomerate firm, such wage adjustments appear relatively more difficult to accomplish. Workers in the ailing division then see themselves as part of the larger conglomerate hierarchy, within which wage rates other than theirs are not falling, and thus resist changes that will worsen their relative position in that hierarchy. By spinning off the ailing division, the conglomerate sacrifices whatever economies of scale and scope that made its ownership attractive in the first place, but in the process may achieve otherwise unattainable concessions on the compensation package. As more and more divisions of conglomerate companies have encountered difficulties in recent years, it has become increasingly common to see such spinoffs take the form of the so-called "leveraged buy out," in which the parent company retains a strong financial stake in the economic future of the division it sheds. If workers took no interest in the incomes of other workers employed by the same firm, it would be difficult to see why spinoffs of this form would make economic sense.[23]

E. *Nonpecuniary Elements of Compensation*

The discussion has focused up to now on the implications of consumption interdependencies for the structure of monetary compensation for coworkers in a competitive labor market. It is clear, however, that interpersonal comparisons go beyond monetary compensation to embrace a broad variety of status-related job characteristics.

The logic of the model summarized in Section I can easily be extended to show, for example, why a firm with several vice presidents might experience a competitive advantage over an otherwise identical firm with only one vice president. To illustrate this point, let us consider an earlier day in which each of numerous identical firms had one president and one vice president. What happens in such an environment if an entrant then establishes a firm with a president, an executive vice president, and vice presidents for both production and marketing? To the extent that the title of vice president confers greater status than the existing titles of marketing or production director, there will be marketing and production directors from incumbent firms willing to supply their services to the entrant at salary levels less than they currently enjoy. But the fundamentally reciprocal nature of status suggests that while the bestowal of these titles confers advantage to their recipients, it simultaneously acts to reduce the status of those subordinate employees whose job descriptions have not changed. Given dispersion of preferences concerning status, however, we know that in large markets there will be those who are willing to pay more for the privilege of having the titles than would be required to compensate subordinates for the loss they suffer as a result of their issuance. We may expect, then, that some successful firms will have multiple vice presidents and will pay these executives less, and their lower staff more, than will firms with only one vice president.

Titles are free for the giving, but many nonsalary compensation items we commonly observe cost the firm real resources to provide. Except for the obvious tax motivations that underlie the provision of certain fringe benefits, why would a firm spend large sums providing special amenities for some employees rather than give them an equivalent salary increment which they could then spend or not as they saw fit on those same (tax deductible) job-related amenities? Why, for example, do chaired professorships in universities carry with them sizable research budgets rather than salaries that are larger by the amount of those budgets? Such nonsalary compensation items may be interpreted as devices for rewarding relatively

[23]An alternative interpretation of the leveraged buy out is that the conglomerate firm has expanded into the region of decreasing returns to scale and scope, and wishes to retrench. This interpretation predicts that we should not see firms that shed divisions in this fashion simultaneously involved in expansions elsewhere. The alternative interpretation offered above, by contrast, is consistent with such expansions.

productive employees without stimulating excessive ripple effects throughout the lower end of the wage scale. If salary levels are the primary focus of coworker interpersonal comparisons, it may well be cheaper, for example, for a university to provide a chaired professor with $60,000 in salary and a $60,000 research budget than to provide a utility-equivalent salary level of, say, $75,000.

IV. Status and Fairness

It is often suggested that "equity considerations" account for why internal wage structures are so much more egalitarian than the ones predicted by the marginal productivity theory of wages. This observation rings true at an intuitive level, but it raises the question of why, if equity is truly what they seek, do the best workers in existing firms not join new firms with other workers who are just as productive as themselves? Such firms could cover their costs by paying each worker his marginal product, and it would be literally impossible to construct a more equitable working environment than that. Is it reasonable to suppose that feelings of altruism induce the best workers to remain part of the heterogeneous work forces of their current firms, thereby cross subsidizing the less productive workers in those firms? Possibly so, but if altruism really were the high-ranked worker's main concern, he could accomplish much more for his money by being paid his marginal product and contributing the difference between that amount and the amount he now earns to an international relief agency.

As an alternative interpretation of why so-called equity considerations may help explain the egalitarian internal wage structures we observe, consider the following: if workers were paid their marginal products in firms composed of workers of heterogeneous ability levels, the result would be that high-ranked members of those firms would be getting something of value free of charge—namely, their high-status positions. At the same time, the low-ranked members of those firms would be in the reciprocal position of giving away something that has value. Getting things of value without paying for them and being asked to provide things of value without being compensated both offend our sense of fairness, as that term is conventionally understood. Given that such outcomes do indeed offend, it then appears to make perfectly good sense to say that a wage structure in which each worker receives his marginal product is an "unfair" wage structure.

It thus appears difficult to divorce common usage of the term fairness from considerations related directly to people's desires to occupy high-ranked positions in the earnings hierarchies to which they belong. Thought of in this way, what people call a "fair" wage structure is one in which people's earnings embody implicit compensating payments that reflect where they stand in their respective earnings hierarchies. These observations suggest that additional insights might be afforded by viewing fairness as a good like any other that is traded in the marketplace, instead of treating it as an abstract notion, best relegated to the political arena for discussion. To the extent that individuals in society do have broad freedom in their choice of coworkers, friends, and neighbors, it will be important to recognize that society's beliefs about fairness are as richly and fully reflected in the structure of private market transactions as they are in political decisions regarding tax and public expenditure policies.

This observation, in turn, suggests that we can learn something about the value people place on equity by observing the premiums they pay or receive in return for occupying various positions in the income hierarchies of the groups in which they work. Additional research along these lines promises high returns, for while equity has always played a major role in political decision making, we remain in the awkward position of having no consistent framework for measuring the value of equity when its advancement conflicts, as it usually does, with progress toward other objectives whose value we can measure.

V. Concluding Remarks

This paper began by posing the question, "are workers paid their marginal products?" For a limited and probably not very representative sample of occupations for which

coherent measures of marginal products could be devised, we may conclude that the answer to this question seems to be no; in each instance examined, the most productive members within an organization appear to be paid substantially less than their marginal products while the least productive members appear to be paid substantially more.[24]

Such findings as these are sharply at odds with the characterization of labor market equilibrium that emerges from traditional neoclassical models in which workers are assumed to take no interest in the incomes of their coworkers. Yet abundant evidence supports the claim that workers *do* care about their relative standing in the income hierarchies of the groups to which they belong. If so, and if each worker is free to chose his coworkers, I have demonstrated that a heterogeneous group of workers cannot coexist under the roof of one firm in the absence of an array of compensating wage differentials that reflect the relative standing of each worker in the income hierarchy of the group. This is indeed the pattern we appear to observe in the cases examined in this study. Once the presence of interdependent preferences is granted, this pattern creates no tension whatever for the economist's traditional characterization of the behavioral objectives of individuals and firms. On the contrary, the failure to observe such a pattern in internal wage structures under these circumstances would itself raise questions that would prove very difficult for traditional theories of the labor market to answer.

[24]Where interdependencies between workers are important, we may, of course, continue to say that wages will equal marginal products, provided we redefine marginal product to include the value of external effects created by the presence of individuals within specific groups. With the definition of marginal product thus broadened, high-ranked employees' claims for payment would be measured not only against the output for which they are directly responsible, but also against whatever claims for payment their salaries provoke from employees of lesser rank. Similarly, the marginal products of a group's least productive members must be broadened to include whatever value their presence creates for higher-ranked members of the group.

REFERENCES

Alchian, Armen A. and Demsetz, Harold, "Production, Information Costs, and Economic Organization," *American Economic Review*, December 1972, *62*, 777–95.

Brown, Wilfred, *Piecework Abandoned*, London: Heineman Educational Books, Ltd., 1962.

Burton, John and Addison, John, "The Institutionalist Analysis of Wage Inflation: A Critical Appraisal," in Ronald Ehrenberg, ed., *Research in Labor Economics*, Greenwich: JAI Press, 1977, 333–76.

Dalton, Melville, "The Industrial 'Rate Buster': A Characterization," in T. Lupton, ed., *Payment Systems*, Middlesex: Penguin Books, 1972, 64–91.

Doeringer, Peter and Piore, Michael, *Internal Labor Markets and Manpower Analysis*, Lexington: D.C. Heath, 1971.

Dunlop, John T., "The Task of Contemporary Wage Theory," in George W. Taylor and Frank C. Pierson, eds., *New Concepts in Wage Determination*, New York: McGraw-Hill, 1957, 127–39.

Easterlin, Richard A., "Does Economic Growth Improve the Human Lot? Some Empirical Evidence," in Paul A. David and Melvin Reder, eds., *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz*, Palo Alto: Stanford University Press, 1973, 89–125.

Festinger, Leon, "A Theory of Social Comparison Processes," *Human Relations*, No. 2, 1954, *7*, 117–40.

Frank, Robert H., "Interdependent Preferences, Coase, and the Competitive Wage Structure," Working Paper No. 19, Center for the Study of American Political Economy, Cornell University, 1983.

______, *Choosing the Right Pond: The Economics and Politics of the Quest for Status*, New York: Oxford University Press, forthcoming 1985.

Freeman, Richard B., "Unionism and the Dispersion of Wages," *Industrial and Labor Relations Review*, October 1980, *34*, 3–23.

Groves, Theodore, "Incentives in Teams," *Econometrica*, July 1973, *41*, 617–31.

Hamermesh, Daniel S., "Interdependence in

the Labour Market," *Economica*, November 1975, *42*, 420–29.

Harris, Milton and Raviv, Artur, "Some Results on Incentive Contracts with Application to Education and Employment, Health Insurance, and Law Enforcement," *American Economic Review*, March 1978, *68*, 20–30.

Hicks, J. R., "Economic Foundations of Wage Policy," *Economic Journal*, September 1955, *65*, 390–404.

Lazear, Edward and Rosen, Sherman, "Rank Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, October 1981, *89*, 841–64.

McKersie, Robert, "Payment by Results Systems in the United States," International Management Seminar on Forms of Wage and Salary Payment for High Productivity, Supplement to Final Report, Versailles, 1967, 215–36.

Mangum, Garth L., *Wage Incentive Systems*, Berkeley: Institute of Industrial Relations, University of California Press, 1964.

Marriott, R., *Incentive Payment Systems: A Review of Research and Opinion*, London: Staples Press, 1957.

Merton, Robert K. and Kitt, Alice S., "Contributions to the Theory of Reference Group Behavior," in Robert K. Merton and Paul F. Lazarsfeld, eds., *Continuities in Social Research: Studies in the Scope and Method of "The American Soldier,"* Glencoe: Free Press, 1950, 40–105.

Mincer, Jacob, "Union Effects: Wages, Turnover, and Job Training," Working Paper No. 808, National Bureau of Economic Research, 1981.

Nisbett, Richard and Ross, Lee, *Human Inference: Strategies and Shortcomings of Social Judgment*, Englewood Cliffs: Prentice Hall, 1980.

Pencavel, John H., "Work Effort, On-The-Job Screening, and Alternative Methods of Remuneration," in Ronald G. Ehrenberg, ed., *Research in Labor Economics: An Annual Compilation of Research*, Greenwich: JAI Press, 1977, 225–58.

Roethlisberger, F. J. and Dickson, W. J., "Management and the Worker," in T. Lupton, ed., *Payment Systems*, Middlesex: Penguin Books, 1972, 21–34.

Rosen, Sherwin, "Authority, Control, and the Distribution of Earnings," *Bell Journal of Economics*, Autumn 1982, *13*, 311–23.

Roy, Donald, "Quota Restriction and Goldbricking in a Machine Shop," in T. Lupton, ed., *Payment Systems*, Middlesex: Penguin Books, 1972, 35–63.

Sayles, Leonard, "The Impact of Incentives on Intergroup Work Relations," in T. Lupton, ed., *Payment Systems*, Middlesex: Penguin Books, 1972, 92–102.

Scully, Gerald W., "Pay and Performance in Major League Baseball," *American Economic Review*, December 1974, *64*, 915–30.

Shavell, Steven, "Risk Sharing and Incentives in the Principal and Agent Relationship," *Bell Journal of Economics*, Spring 1979, *10*, 35–73.

Stiglitz, Joseph E., "Incentives, Risk, and Information," *Bell Journal of Economics*, Autumn 1975, *6*, 552–79.

Streeten, Paul, "Wages, Prices and Productivity," *Kyklos*, No. 3, 1962, *15*, 723–33.

Taylor, S. E. and Fiske, S. T., "Salience, Attention, and Attribution: Top of the Head Phenomena," in L. Berkowitz, ed., *Advances in Experimental Social Psychology*, Vol. 11, New York: Academic Press, 1978.

Thurow, Lester, *Generating Inequality*, New York: Basic Books, 1975.

Tversky, Amos and Kahneman, David, "Availability: A Heuristic for Judging Frequency and Probability," *Cognitive Psychology*, September 1973, *5*, 207–32.

Whyte, William F., *Money and Motivation*, New York: Harper, 1955.

Williams, Robin M., "Relative Deprivation," in Lewis A. Coser, ed., *The Ideal of Social Structure: Papers in Honor of Robert K. Merton*, New York: Harcourt, Brace, Jovanovich, 1975, 355–78.

Wolf, William B., *Wage Incentives as a Managerial Tool*, New York: Columbia University Press, 1957.

National Association of Realtors, *Inside the Real Estate Business*, Washington: Acropolis, 1982.

National Industrial Conference Board, "Incentive Plans for Salesmen," *Studies in Personnel Policy*, No. 217, 1970.

"Contestability" vs. Competition

By WILLIAM G. SHEPHERD*

Recent writings by William Baumol, Elizabeth Bailey, John Panzar, and Robert Willig have presented the case of perfectly free entry and exit, where an entrant can replace any existing firm before that firm can respond.[1] Such ultra-free entry provides efficient outcomes, they suggest, not only in theory but in actual markets. Among the desirable results are said to be zero profits, Ramsey optimal prices, efficient production and market structure, innovation, and an avoidance of cross subsidies in pricing: all this even in pure monopolies.

Martin Weitzman (1983) and Marius Schwartz and Robert Reynolds (1983) have questioned the realism and robustness of these results. But the Baumol-Panzar-Willig reply was that these criticisms only "enrich the analysis and open new areas for pertinent research" (1983, p. 491). A review by Michael Spence (1983) raises only minor doubts, and William Brock (1983) calls the analysis "important" while noting limits on its determinancy and robustness. As further papers about ultra-free entry appear, the topic is being installed not only in the literature, but also in graduate curricula and in policy debates about antitrust and regulation. An assessment now would be timely.

Baumol et al.'s whole analysis is a sizable theoretical accomplishment, combining many elements. The major part of it is an analysis of multiple products and joint costs, which is already gaining wide acceptance. The other part, about ultra-free entry, has been offered with unusual self-praise.[2] Baumol et al. have also drawn many policy lessons from the theory, all suggesting that actions should be directed at entry barriers instead of internal conditions. For example, antitrust actions to reduce market dominance may be "mischievous and antisocial," and regulatory limits on dropping utility services are "perverse" (Baumol, 1982, p. 14).

I try here only to appraise the analysis of ultra-free entry, as a normative contribution to industrial organization. It is largely separable from the multiproduct and sustainability analysis, which I leave for other writers and occasions.[3] Even so, a lengthy discussion is needed, because complex issues are involved. In Section I, I place ultra-free entry in context of the evolving field of industrial organization. Section II assesses the conceptual validity of the Baumol et al. analysis in representing the nature of competition. Section III considers empirical issues in measuring and testing ultra-free entry.

Terminology. At the start, the use of terms needs revision. Baumol et al.'s vague term, "contestability," particularly needs to be replaced by the more accurate phrase, "ultra-

*University of Michigan, Ann Arbor, MI 48109. I am especially grateful to Mark Bagnoli and Barton Lipman for ideas and advice. I have also benefited from discussions with William Adams, William Baumol, Severin Borenstein, Kenneth Boyer, Ralph Bradburd, Daniel Fusfeld, John Heywood, John Kwoka, Takeo Nakao, John Panzar, David Ross, F. M. Scherer, Harry Trebing, Hal Varian, Robert Willig, and seminar participants at Amherst College, the universities of Chicago, Columbia, Doshisha, Louvain, Michigan, Michigan State, Nankai, and Williams College. Research support from Williams College and the University of Michigan is gratefully acknowledged. I remain solely responsible for the paper's content.

[1]The main source is Baumol, Panzar and Willig (1982, cited as B-P-W), which includes a laudatory Foreword by Bailey and whose chapters are adaptations of many individual papers. See also Baumol's summary (1982), Bailey and Baumol's review of cases (1982), and Ioannes Kessides' empirical paper (1982).

[2]To take a few examples, it is "a new theory of industrial organization," which "will transform the field and render it far more applicable to the real world" and be "extraordinarily helpful in the design of public policy" (B-P-W, 1982, pp. xiii, xxii). The invisible hand "seems to rule almost everywhere" that ultra-free entry exists (Baumol, 1982, p. 2). And "we hope to provide a unifying framework for a pure theory of industrial organization where none was available before" because the pieces of the previous literature "are generally disconnected" (B-P-W, 1982, p. 3). The aim is "ambitious" and the contributions are "fundamental" (B-P-W, 1982, pp. vii; 1).

[3]In my (1982b) article, I have offered serious doubts that "sustainability" offers useful policy guidance for regulation.

free entry." Baumol et al. focus on several aspects of entry, but their results hold only for the ultra-free case, which involves the following three conditions.

1) *Entry is free and without limit.* A new firm can do far more than gain a foothold quickly, as conventional free entry envisages. The entrant can immediately duplicate and entirely replace any existing firm, even a complete monopolist. "Entrants can, without restriction, serve the same market demands and use the same productive techniques as those available to the incumbent firms" (B-P-W, 1982, p. 5). I will call this *total entry*. There are no costs or significant lags in entry, and the entrant can match all dimensions of size, technology, costs, product array, brand loyalties, and other advantages of all existing firms.

2) *Entry is absolute.* The entrant can establish itself before an existing firm makes any price response. If the entrant obtains an advantage, even a tiny price difference, it will prevail absolutely and displace the existing firm, with no interaction or sequence of moves. This Bertrand-Nash assumption of no response holds even if a pure monopolist faces elimination and would have to abandon its monopoly pricing entirely.

3) *Entry is perfectly reversible.* Exit is perfectly free, at no sacrifice of any cost. Sunk cost is zero.

These conditions are pure, and the deductive results hold only when they hold. Under any departures from the pure conditions, Baumol et al.'s deductive analysis becomes speculative. One can revert then to the extensive literature on entry barriers for guidance in estimating the outcomes.

Yet "contestability" seems to refer only to those *imperfect* entry cases, rather than to the pure case. The word is informal and intuitive, implying interactions and imperfections that the ultra-free entry model rules out. By dictionary definition, a *contest* is a struggle, whose uncertain outcome depends on a series of actions and reactions.[4] Virtually every market is at least partly vulnerable to entry; that is "contest-*able*." Yet ultra-free entry permits *no* contest; the entrant either does or does not prevail over the incumbent, immediately and absolutely. Baumol et al.'s own usage displays the problem, frequently slipping between the pure meaning and the English language meaning. That is crucial, because their result holds only for the pure case. Part of the seeming generality of Baumol et al.'s conclusions arises from the word's vagueness and ambiguity. The premier question is whether the ultra-free entry results apply when entry is not ultra free. The term contestability confuses precisely that issue.

Moreover, Baumol et al.'s stress is on the prevention of entry by anticipatory price restraint, so that actual entry need never occur. Therefore the deterrence of entry is crucial, not a possible interaction (a "contest") *after* entry.

So the term contestability has two defects. It misplaces attention from entry conditions to a nonexistent post-entry struggle. And it implies degrees of variation in entry that are not logically permissible. The phrase *ultra-free entry* more precisely reflects the crucial conditions. I will use it here, and I urge its general adoption. Continued use of "contestability" will encourage needless confusion between true ultra-free entry and all of the many other entry conditions.

I. The Setting

To understand what Baumol et al. have added to the field of industrial organization, one needs first to review what was already there. The brief summary in this section is needed because ultra-free entry has been presented *en bloc* as a wholly new system, with little discussion of its setting and predecessors. The field has deep roots, sophisticated content, and complex trends.[5] Most of its basic concepts have long been familiar, and

[4]The verb *contest* means "to make the subject of dispute, contention or battle"; the noun *contest* is "an earnest struggle for superiority or victory." *Contestable* is "capable of being contested" (see Webster's *Third New International Dictionary*, 1981).

[5]For the main lines, one may begin with Joe Bain (1956; 1968), George Stigler (1968), William Fellner (1949), Oliver Williamson (1975); and such textbooks as F. M. Scherer (1980), Alfred Kahn (1971), John Blair (1972), and myself (1979). B-P-W give little sign of incorporating the substance of these writings and the underlying literature.

they have been subjected to increasingly thorough empirical tests.

A. *Core Topics*

Industrial organization is about the nature of the competitive process and the effects of monopoly distortions (and possible benefits) in the variety of actual markets. *Real* conditions are the ultimate guide in sifting the many concepts that theorists can invent. Appraising ultra-free entry is one more episode in the century-old process of assessing competing ideas.

The field has evolved since Marshall's time from the neoclassical analysis of polar competition and monopoly. Those border cases, based on the firm's own demand (approximately horizontal for competition, downsloping for monopoly), were soon clarified.[6] The modern field of industrial organization has been occupied mainly with the complex intermediate gradations of partial monopoly, tight and loose oligopoly, entry, vertical integration, and so on.

A basic distinction is between internal market conditions and external conditions. *Internal conditions embody the degrees of actual competition and monopoly* among firms already existing inside the market. These internal conditions of actual competition include both structure and behavior. Such features as market structure, demand inelasticities, lags, brand loyalties, price discrimination, cost differences, information gaps, and strategic behavior have attracted a vast volume and variety of research.

External conditions of potential competition, from firms outside the market who may try to enter in the future, may also influence the choices by firms inside the market. Such potential competition could even supersede actual competition, in certain situations. But will that generally, or always, occur? The question has puzzled specialists for nearly three decades, and Baumol et al.'s discussion poses it starkly.

B. *Trends*

In the early years of the field in 1880–1910, internal conditions were the main concern. As mergers created dominant firms with market shares above 60 percent in scores of markets, market share was seen (as in past centuries) to be the main element of market power. Though potential entry was occasionally mentioned (see John Bates Clark, 1907), it was seen as a peripheral matter.

Many industries then evolved toward oligopoly during 1910–30, and the landmark studies in 1933 by Edward Chamberlin and Joan Robinson quickly made interactions among oligopolists the new popular topic. By 1939 the first complete sets of four-firm concentration ratios became available, giving research on oligopoly a valuable (if imperfect) empirical basis. Soon the von Neumann-Morgenstern (1944) theory of games stirred further excitement over the "oligopoly question." So encouraged, the explorations of the indeterminancies of joint maximizing by oligopolists eclipsed the study of individual-firm market power. Excess profit now seemed to arise mainly from the collusion of oligopolists, if they held more than a threshold level of tight-oligopoly concentration, rather than from individual firms exerting their own market power. The differences among individual firms came to be neglected, in the rising preoccupation with oligopoly and concentration ratios. Market shares and price discrimination came to be ignored.

In 1956, Joe Bain added the concept of entry barriers, as an element that could modify internal market power. The logical possibility that free entry could neutralize internal market power was known from the outset, but it was set aside as an extreme case. Both barriers and entry have been controversial, but they are known to contain several elements and possible sources. About ten main alternative sources of entry barriers have now been discussed.[7] The act of entry, a conversion of potential into actual competitors, was

[6]See Richard Schmalansee (1982) for a lucid discussion of demand curves and market power.

[7]Scherer (1980), William Comanor and Thomas Wilson (1979), and John Heywood (1982) discuss the various elements.

seen as quite separate from the barrier conditions themselves. Carrying the idea of barriers to still other dimensions, Richard Caves and Michael Porter have added barriers to *exit* (1976) and then barriers to *mobility among parts of a market* (1977). Analysts vary in the weight they assign to entry barriers; most still recognize internal conditions as primary, but some would give barriers primacy in some situations.

Scale economies were also studied by Bain, Leonard Weiss, Scherer, and many others after 1940. The effort is to determine how much concentration is dictated by economies. Structure is treated as endogenous, in extensive research. In fact, most average costs appear to fit an L-shaped curve, not a U shape. The technical economies of scale explain only part of concentration. There is much excess market share (the increment of actual market share over minimum efficient scale), as is shown by Scherer (1980, pp. 91–119) and myself (1982c).

Three other strands in the field since 1960 have been important. *One* is the rise since 1957 of the optimistic new-Chicago-school view, which regards all elements of market power (internal and external) as small and/or short-lived. Any existing market power is held to reflect economies of scale; therefore it is justified by efficiency.[8] A *second* strand (since 1968) is a reassertion of the importance of firms' individual market shares.[9] Market share appears to be the central element of market structure. It is closely associated with excess rates of return, and it is the direct measure of competitive impacts upon the firm. For leading firms, losses of market share (to large, small, or new competitors) are the crucial threat. But such losses are usually incremental and slow. In parallel to the renewed focus on market shares has come a revival of interest in price discrimination, which is a single-firm activity, not a matter of oligopoly coordination (see Scherer, 1980, ch. 11).

The *third* strand has been the advent of applied theorists to industrial organization in the 1970's. Their modelling has yielded a number of insights, especially about game theoretic choices by duopolists. Entry is commonly presented as a two-firm game between the incumbent and a potential entrant. In the process, the distinction between existing competitors and new ones (between internal and external conditions) is often lost. The incumbent firm is generally assumed to be passive to the newcomer, a view which reverses reality but permits the derivation of determinate solutions. Ironically this flowering of duopoly theory has occurred while actual tight oligopoly in the U.S. economy has shrunk by half (see my 1982a article). And equal-firm duopoly scarcely exists anywhere in real markets.

C. *Ultra-Free Entry*

Entering this changing scene after 1977, Baumol et al. have taken some of the trends to the extreme: external conditions are assumed to dominate internal conditions; exit is elevated to be a crucial element; the models are pure Bertrand-Nash, with perfectly passive incumbent firms; and Baumol et al.'s optimism about efficiency appears to exceed even Chicago school levels.[10]

The normative lessons are based solely on maximizing static consumer's and producer's surplus, as is common in recent modelling. Innovation, equity, and other social values are held aside, even though those values may exceed and offset the static results. The analysis therefore offers only a first step. Multiple-product conditions are also explored at length, in pure theory, but the claims made for it are realistic, even understated. In contrast, the ultra-free entry analysis is quite brief compared to the large lessons which Baumol et al. draw from it.

Ultra-free entry, with a focus on exit, is indeed new, but only as it goes beyond the previous versions of free entry. In fact, Caves

[8]See Richard Posner's article (1979); also Robert Bork (1978) and Sam Peltzman (1977).

[9]See my (1972, 1975, 1979, 1982c) articles, John Kwoka (1979), David Ravenscraft (1983), and Stephen Martin (1983).

[10]In personal discussions, Baumol et al. do not admit to this degree of optimism, but their writings suggest it clearly.

and Porter (1976) had explored exit barriers more fully. In the Baumol et al. approach, an entrant contemplates not only the ease of entering the market, but also the cost of leaving it (of reversing its entry). Free entry is necessary, but it is now not sufficient, to yield Pareto-efficient results. Free exit is also necessary and (given free entry) it is sufficient.[11] The sources of exit barriers are defined simply as sunk cost; zero sunk cost provides perfectly free exit (occasionally mentioned as costlessly reversible entry).

Baumol et al.'s one significant industrial example of ultra-free entry has been the airline industry. Each existing or new firm can send aircraft anew into city-pair routes where existing fares are significantly above cost. Exit has low costs, so that even a brief entry may be profitable. The threat of such entry may hold price down to cost, throughout the hundreds of city-pair routes, even those that are seemingly monopolized by one airline.

It is easy to reject ultra-free entry as a kind of fantasy whose key elements contradict reality. Moreover, Baumol et al. often are unclear whether ultra-free entry displaces the "older" industrial organization or merely supplements it with "insights." Is this a major development or a curiosum? A careful appraisal is needed, invoking the cumulation of research into industrial organization. One cannot endeavor in this short space to treat the many issues raised by Baumol et al.'s large book and many articles. But certain main points stand out, and I will present them one by one. The discussion is on two levels; there are seven matters of concept in Section II and seven issues of measurement in Section III.

II. Conceptual Issues

How consistent are the assumptions of ultra-free entry and how general are the lessons? The following points are primarily matters of logic, although some empirical elements creep in.

1. *Ultra-free entry involves inconsistent assumptions*. By assuming that total entry, profitable selling and exit can occur before the existing firm responds, Baumol et al. obtain determinacy. Moreover, the prospect of such brief ("hit-and-run") but total entry forces even a pure monopolist to set prices at efficient levels. Such an *entry-nullified monopoly* is Baumol et al.'s distinctive result.

Yet the assumptions are contradictory. If entry is sufficiently trivial, it may indeed avoid a response. As Baumol-Panzer-Willig put it: "However, if an entrant's output is 'small' relative to that of the industry, the magnitude of these required adjustments may be 'small', and hence it may be justifiable for the entrant to ignore them" (1982, p. 11). All elements are trivially small. But ultra-free entry also assumes *total* entry. The two assumptions, of trivial entry and total entry, are opposites. If entry is trivial, it has no force. If it is total (or even merely significant), then the no-response assumption is not tenable.

If the Bertrand-Nash nonresponse assumption holds, then the model does not admit significant entry, which can influence the incumbent. Total entry, which would entirely duplicate and replace even a monopolist, would be particularly absurd in a Bertrand-Nash model. Yet that case is what ultra-free entry assumes, and where its special superiority over competition in the market is said to occur. In short, the Bertrand-Nash basis appears to rule out significant normative conclusions.[12]

[11]At some points, Baumol et al. appear to regard exit barriers as just part of entry conditions, important only as they raise entry barriers. But mostly the writings stress exit as a major new element, which is separable from entry barriers and can be important even if entry barriers are not.

[12]Baumol has suggested that, instead, entrants can offer long-term contracts as a plausible way to obtain total entry before response can occur. Entry occurs, as it were, with a stroke of a pen rather than by the physical creation of the entrant's capacity. Indeed the contract itself could apparently become a tradable item with a market value equal to the incumbent's possible rents. But this suggestion is too easy. It merely transfers the action to contracting wars, and it assumes that incumbents are unable to wield any advantages in offering their own long-term contracts. The same contradictory assumptions still apply, but now in a slightly different setting. Indeed, the incumbent could set standing offers to beat all credible offers by entrants, thereby rendering them not credible and therefore weightless.

It also appears to exclude Baumol et al.'s aspiration to make market structure fully endogenous under ultra-free entry, so that structure is driven to the optimal configuration. Because "small" entry exerts little influence by definition, it could not be assumed to enforce any specific structure. Moreover, the normative objective is to prevent *excess market share* when there are constant costs above minimum efficient scale. "Small" entry does not appear to assure that.

We are left with three cases of entry. 1) First is *ultra-free entry*, with its inconsistent assumptions of total entry and trivially small entry. Since one can't have it both ways, the analyst must choose between the assumptions. Either there is 2) *minimal* entry: trivially small, with no response, or 3) *strong* entry, on a large scale (up to total entry) but drawing strong resistance. Both minimal and strong entry are mixed cases, involving gradients and interactions. Neither case provides clear normative results. Indeed, both cases are covered in the previous literature.

2. *Even if it were consistent, ultra-free entry lacks generality, because of its extreme character*. Total entry, without a timely response by the victim, is assumed to occur even from a tiny price difference, which the incumbent could easily match or retaliate against. There seems to be no general theory here, only an odd special case. In contrast, pure competition among price-taking firms rests on plausible, consistent assumptions, even in the extreme case of hundreds of firms. To establish ultra-free entry as a general case, one will probably need to show that its normative results are robust even to large departures from the assumptions, that important real markets are open to ultra-free entry, and that many entry episodes combine minimal price differences with maximum impacts. Research to that end has scarcely begun.

3. *The analysis relating costs and entry-exit barriers is imaginative, but it raises three important doubts*. First, fixed costs are claimed (or defined) to be irrelevant to entry barriers. But that is true only if the entrant can match the incumbent firm's output completely, by means of total entry. Otherwise the fixed cost per unit is higher for the entrant (with its smaller output), thereby permitting the incumbent to maintain a price differential without inducing entry; in short, an entry barrier exists.

So total entry must be presumed. Only in that extreme case is the Baumol et al. cost analysis, which tries to separate fixed cost from sunk cost, fully valid. Total entry is an extreme (and inconsistent) assumption, as I have noted, and the analysis does not extend beyond it.

Second, sunk costs are closely, and inversely, related to time duration. In any market, sunk costs are larger in the short run than during longer periods, by definition. Therefore sunk costs are most likely to be highest, and to cause entry barriers, precisely in those very short periods when the incumbent is assumed *not* to respond to entry.[13] Sunk cost as the cause of entry barriers makes it particularly implausible that ultra-free entry can occur. The focus on sunk cost therefore introduces another logical problem about the consistency of the model of ultra-free entry.

Third, virtually all production requires specific assets which cannot be transferred or sold costlessly. This applies to physical equipment, advertising, *R&D*, expert skills, and the other commitments needed to establish entry. Fixed and sunk costs commonly overlap and are sizable, as Spence (1983) and others note. Zero sunk cost is therefore a doubtful, counterfactual assumption for a general theory.

4. *The assumption that external conditions dominate internal ones is eccentric.* If entry is

[13] More generally, the issue is whether *time lags* or the *height* of barriers is critical. Baumol, Panzar, and Willig (1982) stress sunk cost (that is, exit-barrier height); Weitzman, Schwartz and Reynolds, and Spence have stressed the time lags. In fact, both are involved; height and lags are related, to be defined in terms of each other. For exit barriers, a longer lag reduces sunk costs by converting some of them into variable costs. Entry-barrier height is definable as the time period required for an entrant to achieve a given market share. This unity between barrier height and time lags eases the controversy, but not the problems of defining actual barriers, using arbitrary time periods.

ultra free, then potential competition effaces actual monopoly by definition, as a tautology.

This pure result can be illustrated by a kinked demand curve for the incumbent firm. Demand is horizontal to the left of the kink, at the efficient price-cost level. If entry is less than ultra free, that segment of the demand curve has some slope, and the determinate normative results are lost. The familiar analysis of barriers then takes over, where entry is largely secondary and blends in with other marginal competitive factors. Whenever there is more than one incumbent firm, each has to consider its actual competitor(s) as an immediate threat to its market share. Even if an existing rival is small, it is present, already exerting the pressure that a potential entrant might have *if* it chooses to enter. In theory, an incumbent firm might ignore actual competitors, directing its attention and policies exclusively toward a possible addition to the actual competitors. There may be small actual markets which are intimidated by a very large firm, which is threatening to enter on a big scale. But to assume that condition as the general case is farfetched and wrong for normal markets.

5. *Structure is not more endogenous than in previous analysis.* As before, Baumol et al.'s analysis holds for the case of ultra-free entry. It posits U-shaped cost curves with single lowest-cost points. If one could ignore the conflict of assumptions noted earlier, then ultra-free entry might enforce the efficient number of firms, all at optimal scale.

Pure (or reasonably effective) competition has also yielded this result, despite Baumol et al.'s denial. Competition enforces minimum-cost production. Indeed the claim that in "older" research, structure was "determined exogenously in a manner totally unspecified" (Baumol, 1982, p. 7) suggests an unfamiliarity with the literature. As noted in Section I, the L-shaped average cost curve has come to be regarded as common since 1960, with significant ranges of constant costs. Extensive research has sought to show how scale economies do influence structure; in short, to make structure endogenous to a full analysis. It has also tried to estimate the excess market shares held by leading firms in some major markets. The endogeneity offered by Baumol et al. is simply a possible corollary of the ultra-free entry case. But if entry is not total *and* nonresponded, endogeneity may not occur even in this one case. Baumol offers a "guess" that when entry is less than ultra free, the outcome will give "reasonable approximations to the efficient structures" (1983, pp. 7–8). But that is simply a conjecture, as he notes. And it ignores the forces of *actual* competition inside the market.

6. *The case of a monopoly or dominant firm facing ultra-free entry is probably unstable and transient, not an equilibrium state.* If costs provide for a natural monopoly, the resulting firm will create entry barriers, threaten retaliation, and generally violate the assumptions of ultra-free entry. If costs provide for a degree of natural competition, then entry will probably occur and internal structure will evolve toward competition. The case of an entry-nullified monopoly or dominant firm is unlikely to last even when it does occur.

7. *The Baumol et al. analysis appears to restrict the invisible hand, not broaden its reach.* Baumol et al. suggest that they have broadened the likely scope of competitive results. For example, performance under ultra-free entry is "really outstanding" and this setting is "more ideal than perfect competition" (Bailey and Baumol, 1982). The main possible extension of Pareto outcomes is actually to the short run. If ultra-free entry is valid, then it provides efficiency quickly, rather than after long-run adjustments occur.

Yet ultra-free entry imposes an extra requirement, beyond total entry and nonresponding incumbents. In addition, *exit* must also be perfectly free, with zero sunk cost. Even if entry barriers are absent, exit barriers may exist and prevent the efficient outcome. This new requirement is strict: *any* sunk cost (not just an unusual degree of it) impedes exit.

A corollary is that exit barriers are irrelevant for the market outcome except *when they are higher than entry barriers.* If entry barriers are higher, then exit barriers do not matter. In that sense, the large past literature

on entry has not ignored exit barriers, but rather assumed them to be comparable or lower than entry barriers. That is a reasonable general assumption. Yet Baumol et al. may be correct in emphasizing exit barriers as an important separate element. The issue now appears to be an empirical one. The task is to measure exit barriers, compare them with entry barriers, and show that important low-entry-barrier and high-exit-barrier cases exist. Such findings would give support to Baumol et al.'s new lessons.

Together the three extreme assumptions (total entry, nonresponse, and exit barriers as an element) leave ultra-free entry as a special case, not a general basis.

More broadly, the assumptions and normative conclusions of the Baumol et al. analysis are doubtful on conceptual grounds. Only the ultra-free entry case is new, and it contains contradictions and is narrow.

III. Empirical Issues

The next step of the evaluation is empirical. What factual support is there for the assumptions and conclusions of the model? Is the case important and general? There are many issues, some of them quite complex. Again it seems best to present a series of specific points.

1. *Theory, Facts, and Scientific Methods.* As Baumol et al. concede, there is little empirical basis so far for their theory. One industry, airlines, is presented as a case of ultra-free entry, but its support is doubtful, as I will note below in point 7. Baumol, Panzar, and Willig (1982) offer in chapter 16 some discussion and "checklists" of items to cover in future research, but the ideas are vague and there is no systematic or practical research design.[14] One econometric study of sunk cost (see Ioannes Kessides, 1982) has been prepared more recently; it is suggestive at best (see below, point 5).

Theorizing before testing is often an acceptable approach, but Baumol et al. have taken it beyond reasonable bounds of valid scientific method. Implausible assumptions have been applied on an abstract plane to reach not only "insights," but also emphatic conclusions and wide policy lessons. The system hangs in the air, lacking a foundation or even plausibility. If the adjacent technical analysis of multiproduct conditions were less formidable and the authors less famous, these ideas and claims would seem naive and premature. Because the whole set of writings is so audacious and complex, mingling pure theory with claims about reality, a full evaluation is difficult, and this paper can be only a beginning. In the meantime, ultra-free entry is a theory urgently in search of facts.

Fortunately, the existing field already provides much evidence about the foundations of ultra-free entry analysis, so that even the following brief summary can be suggestive.

2. *Do External Conditions Dominate?* A large inventory of research findings, accrued over many decades, has shown internal conditions to be the primary force.[15] Market shares are commonly the most important element. Repeated testing shows that market shares correlate closely with rates of return and explain much of their variation.[16] If entry conditions dominated, that correlation would

[14]On the crucial question of measuring contestability, B-P-W are quite vague. Here is their entire research discussion of the question:

Determination of the structural contestability of a market requires evaluation of the costs of entry and exit and of the magnitude of unavoidable sunk costs. In particular, the availability of resale markets for durable inputs and their usability in other activities (fungibility) must be investigated since, clearly, the less the financial loss incurred in such a transfer, the lower will be the costs that are truly sunk and the smaller will be the costs of exit. It will also be essential to have a description of the sources of entry and exit costs as well as of the sunk costs to determine whether these costs can readily be reduced by appropriate policy measures. For example, if sunk costs are increased by legal inhibitions to the formation of resale markets, appropriate remedial measures may suggest themselves. Finally, it will be desirable to determine how entry costs are affected by the size of the potential entrant, that is, to estimate the $E(y)$ function roughly. [p. 469]

All of these are complex phenomena, needing precise, thorough discussion and tests.

[15]See Scherer (1980), myself (1979), and Blair, inter alia.

[16]See especially my 1972, 1975, and 1979 articles and numerous studies cited there.

not occur.[17] Firms with high market shares usually have a substantial degree of market power. Conversely, firms with low market shares do not have significant market power, even if they are in the same market as a dominant firm. That condition applies also to new entrants: entry on a small scale usually applies little competitive pressure.

Entry barriers have been shown to have some effects, perhaps reinforcing or interacting with, market structure.[18] But no significant evidence exists that free entry has or will fully neutralize market dominance, much less pure monopoly. On the contrary, there is widespread evidence that dominant firms and monopolies take strong steps to create entry barriers and to retaliate against entry. In fact, no study has thoroughly separated entry barriers from internal structure. Since entry takes effect in stages, primarily by reducing existing market shares, the entry research tends to affirm the importance of internal conditions.

One can also obtain valuable guidance on the issue from the basic patterns of business experience and policy. There is a pervasive concern with market shares, throughout industrial life and the business press. The struggle for higher market shares (as the proximate basis for higher profitability) is intense and nearly universal. In contrast, entry is seen as an occasional matter, rarely crucial and often insignificant compared to the struggles among actual competitors. Entry barriers are rarely discussed as such, or framed as a basic objective. Entry itself usually occurs as a process, growing over time from small footholds. Exit barriers are a minor and unfamiliar possibility, rarely mentioned.

Research and experience therefore confirm the expected dominance of internal over external conditions. Entry is usually a secondary factor. This places the burden of proof on Baumol et al. to show that entry is actually a dominant factor.

Moreover, internal and external factors are correlated directly, not inversely (see Scherer, 1980, ch. 3). The two conditions interact; high barriers breed high market shares and concentration; and vice versa. Also, potential entrants are numerous mainly when actual competitors are also numerous, in competitive industries with small scale economies. As a result, there are probably few significant markets combining high internal monopoly with low entry barriers. Indeed the total share of pure-monopoly and dominant-firm markets in the U.S. economy is only about 5 percent of *GNP* (see my 1982a article). If cases of entry-nullified monopoly or dominant firms are only a small share of those cases, they will be an insignificant share in the whole economy.

3. *Extreme Assumptions and a Lack of Cases.* Research has provided no cases of ultra-free entry into markets where incumbents hold substantial market shares. Reasonably free entry, in a loose sense of low entry barriers, does exist for many markets, perhaps accounting for over one-quarter of *GNP*. Much of agriculture, various services and trade, airlines, brokerage, and some import-affected industries are the main instances (see my 1982a article). But, of course, almost all of them also have a competitive *internal* structure. Entry therefore has little effect or role.

Consider the attributes of ultra-free entry, one by one. *Total entry* and *nonresponded entry* on a significant scale are both unknown in markets with substantial market shares. In general, entry is slow and occurs in a sequence involving first a foothold and then later expansion. Significant entry virtually always draws strong quick retaliation. The strength and speed of retaliation vary directly with both the incumbent's and the entrant's market shares.

Deliberate *brief entry* (hit-and-run entry) is also rare and unproven. Similarly, *free exit* (zero sunk cost) is also virtually unknown in significant markets with substantial market shares.[19] Sunk costs obviously vary with the

[17] With one exception: scale economies could cause it. But they are known to be insufficient to explain the profit variation (see Scherer, 1980, and myself, 1982c).

[18] Product differentiation has emerged from empirical tests as the most important cause of barriers (Comanor and Wilson; Heywood).

[19] Such costs include many categories besides physical capital, such as *R&D*, advertising to establish brand loyalty, and training to create special workers' skills. These intangible forms are often more fully "sunk" than physical capital, which can be leased or resold.

firm's market share, so trivial firms can exit at little absolute cost. But significant ones cannot.

Accordingly, no cases of entry-nullified monopolies or dominant firms have been established. These categories appear to be empty. Many dominant firms have declined, of course, but they do not reflect ultra-free entry. The significant declining cases since 1960 reflect antitrust actions, rising imports, and incursions by smaller firms.[20] Entry has played some role, but it has been secondary and not ultra free; and imports have not taken large market shares swiftly.[21]

Two promising cases—airlines and long-distance telephone service—do not display ultra-free entry. Airline shifts have not been instantaneous nor nonresponded (see more below); and entrants into long-distance telephone service gained only 5 percent of the market in four years, despite regulatory limits on AT&T's ability to respond.[22] Many other dominant firms have developed high barriers to entry, protecting high market shares and high profit rates.

Finally, no important cases have been identified which have exit barriers "higher" than, or separate from, entry barriers. The only important exit barriers are artificial ones, such as regulatory rules against abandoning railroad service. Therefore the analysis of sunk cost, as if the exit barriers it creates were an important natural feature of markets, is a detour. If exit barriers are artificial, then they are no part of a new general theory of industrial structure. We are left again with the familiar research on entry barriers. That research is, of course, often unclear, for the following reasons.

4. *Barriers, Entry, and Market Definitions are Unclear*. The ensuing complications do not disprove Baumol et al.'s approach. But they do make the subject more complex than their writings recognize. Scientific support is hard to develop, and added caution about ultra-free entry seems advisable.

Defining Markets. The ultra-free entry analysis has assumed that markets are well defined, so that barriers are a clear phenomenon at the edge and new entry is a simple, definite act. Most markets are, in fact, difficult to define. The resulting debates about methodologies and actual markets have long been a central topic in the field of industrial organization (Scherer, 1980, chs. 3, 20; my 1979 book, ch. 9). This problem reduces the clarity and force of entry. Often one cannot tell whether certain firms are inside the market or not (i.e., whether they are actual or potential competitors).

Entry Barriers. Barriers are complex. Bain (1956) noted four main sources of barriers (product differentiation, economies of scale, cost advantages, and specific devices such as patents and crucial inputs). Many other causes have been added since then, including strategic pricing by incumbents, intensive *R&D*, excess capacity, diversification, vertical integration, and sales networks; others can be expected. They may be present singly or in combination; whether they are additive, nonadditive, or multiplicative is still an open question. For example, if a market has three of these sources, does it have three barriers or one barrier; if one barrier, is it three times as high? Do some sources interact, but not others?

Some barrier sources (such as strategic pricing and excess capacity) are discretionary to the incumbent firms. Being *endogenous*, they reflect the internal market power of the firms rather than external barriers. Economies of scale have always been particularly ambiguous in the literature. They can be determinants of structure, part of the normative evaluation of high market shares. Yet they may also be part of structure itself, as a barrier to entry.

These problems have existed all along, causing barriers to be defined loosely, measured imprecisely, and interpreted in conflicting ways. The sharp-edged assumptions of Baumol et al. do not reflect these complexities.

Exit barriers share these problems. Sunk costs resist accurate measurement. Physical capital is subject to varying measures, and its

[20]See my (1982a) article for an evaluation of the influences.

[21]The slowness of import incursions has been noted by Scherer (1982).

[22]No serious observer suggests that MCI or Sprint could quickly (or even eventually) displace AT&T entirely in long-distance telephone service.

sunk component is often unknown (in part because it varies with the time interval involved). Nonphysical forms of sunk cost may often be larger, but they too are difficult to measure. A final problem is setting the criterion for "zero" sunk cost. Taken literally, zero sunk cost cannot occur for a significant market in any brief time period.[23] A "reasonable" or "virtual" basis for zero sunk cost must be defined, if free entry is to have practical meaning. Yet that basis has not been defined, and there may be no clear criteria for doing so.

Entry. Entry is a process separate from entry barriers, and it is also complex. It has at least three elements: size, speed, and probability of occurrence. Also, entry is a *net* amount: gross entry minus exit during each period.

The most general form is continuous entry, starting at infinitely small expected values. As size, speed, and/or probability rise, the expected degree of entry rises continuously. Only when all three elements are substantial is entry significant.

In practice, expected entry is usually continuous, because it occurs by stages and degrees. Each entrant will begin at its preferred scale and then expand as and if conditions are favorable. Most will enter at a small size, in order to minimize risk and permit learning and growth; large entry at full size is unusual. The main impact on the market comes with post-entry growth, but that is *actual competition*.

Measuring entry and exit is still a primitive art. Entry's speed and net size and numbers need to be known. Given this variety, some form of complex comprehensive measure of entry is needed, but none yet exists, either in concept or fact. Merely counting firms is not reliable, because of size disparities and the fact that mergers (which reduce numbers) are not equivalent to exit. As for measuring exit, the relevant scale has not yet been defined: numbers, market shares, and rates of speed need to be included, but in what weighting?

These complexities have kept the nature of barriers and entry from being clearly known and tested. They suggest skepticism about models which employ simple, extreme ideas of these phenomena.

5. *Testing for Sunk Cost's Role*. Even if entry could supersede all internal imperfections, actual conditions will usually deviate from ultra-free entry, as Baumol et al. note. Robustness is then crucial: what gradients apply? Do small deviations result in large departures from efficiency?

Extensive past research has shown some price-cost and profit-rate gradients on internal structure and entry barriers, especially between profit rates and market shares. But empirical tests of ultra-free entry are difficult, because clear definitions and data for the key variables are lacking. Baumol et al. offer virtually no tests and no concrete research design. The one quantitative study is by Kessides, which is cited prominently by Bailey and Baumol (1982) and Baumol, Panzar, and Willig (1983).[24] Kessides essays to test "the critical role which sunk costs ... play in impeding entry" (p. 1). He endeavors to explain *entry* (the change in total number of firms) by *exit*-barriers height, using this basic form:

$$(1) \quad \textit{Rate of entry} = a_0 + a_1\,\textit{Growth} + a_2\,\textit{Scale of Entry} + a_3\,\textit{Initial-Year Profitability} + a_4\,\textit{Sunk Cost}.$$

(Bain objected to such tests, because both entry and barriers are related to pricing.) In Kessides' model, the following reasoning is suggested: growth promotes entry; a larger required scale of entry will discourage entry; high initial-year profitability will encourage

[23]Note by comparison that the extreme assumption that pure competitors are "price takers," with horizontal demand, is plausible and valid for many real markets. On that point, established theory of competition is superior to ultra-free entry theory.

[24]Because it is the only empirical study referenced and is stressed by Baumol et al., I have to consider it at length. It would be preferable to await further testing and debate.

entry; and high sunk costs will discourage entry by raising the cost of an unsuccessful entry bid, followed by exit. Using data from 266 4-digit industries for 1972 and 1977, Kessides reports (mainly) significant coefficients with the expected signs. That is "very encouraging" evidence that the risk arising from sunk costs "constitutes a strong barrier to entry" (p. 36).

Yet the test is indirect and of doubtful meaning. Baumol et al.'s analysis is really about the effect of sunk cost on *price behavior*: they believe that high sunk costs limit exit and entry, permitting high price-cost margins. The actual amount of entry is merely one possible side effect of this causal system, *not* the test of its existence. Indeed, they emphasize that the process works best precisely when *no* entry occurs, because price is restrained in anticipation.

The proper model therefore would explain price-cost outcomes, possibly with this form:

$$\text{(2)} \quad \textit{Profitability} = b_0 + b_1 \textit{ Sunk Cost} + b_2 \textit{ Growth} + \textit{Other Possible Factors}.$$

Tests would be confined to free-entry markets and focused tightly on low ranges of sunk cost, so as to test robustness with small deviations from ultra-free entry. Growth and other terms could be added as filter variables for general market conditions.

Instead, Kessides' model has both profitability and sunk cost as independent variables; profitability is probably endogenous.[25] Also, certain technical data problems are serious.[26]

Kessides correctly incorporates in his model the fact that entry barriers may arise from the incumbent company's actions, and that the rate of entry may vary inversely with concentration in the market. In doing that, he recognizes that internal conditions may govern entry conditions from "outside." But because these "external" conditions are not fully exogeneous, that contradicts the logic of ultra-free entry.

In any event, Kessides' tests do not deal with the border conditions of small deviations from ultra-free entry. Even if sunk cost does relate to entry, the estimated gradient reflects the entire range of sunk cost values. The robustness of the ultra-free entry conclusions to small sunk costs remain unknown.

Kessides analyzes *industrywide* averages, reflecting the assumption that all incumbent firms hold identical market power, under constraint by entry conditions. A natural improvement is to explore *individual company* data as well. Then one can test more accurately the role of sunk costs in the presence of other elements and degrees of competition.[27]

[25]Moreover, profitability and sunk cost are probably collinear. That collinearity is accented because his practical measure of profitability (the price-cost margin, *PCM*) contains depreciation, and so those margins correlate spuriously with sunk costs. If these data are used to estimate equation (2), the conventional correction for the bias in *PCM* is to add a filter variable for the capital-output ratio. But that capital intensity will be tightly covariant with sunk cost, and so the independent role of sunk cost would be impossible to test.

[26]The variable for entry (the simple change in total number of firms) is especially weak. In census industries there is usually a large number of small firms in the fringe. The recorded number of such firms often changes sharply, even though there has been little or no real change in the degree of competition among the main firms in the market. Moreover, census industries are often poorly fitted to true markets. Therefore both the numbers of firms and the raw concentration ratios reported by the census are often highly inaccurate (myself, 1970, 1979; Scherer, 1980). Only *physical* capital is used as a measure of sunk cost; all nonphysical sunk costs are omitted. Kessides uses only machinery as sunk cost, but buildings are probably equally use-specific, and so total physical capital should be included in sunk cost. Census data for capital values are notoriously weak, because they reflect book values subject to varying accounting practices. Price-cost margins are a faulty measure of profitability because of their crudity and the weaknesses in the underlying data. At the least, they need correcting for variations in capital intensity, as has routinely been done in past research.

[27]Several existing models can be easily adapted for such tests (for example, see my 1972 article). Indeed, the use of an asset size variable has already provided a plausible sunk cost estimator; each firm's own asset size measures the (possibly) sunk cost required by a small rival or entrant in order to match that firm's own competitive capacity. If that basis is valid, then asset size should vary directly with rate of return. Alternatively, assets reflect the firm's *own* sunk cost, which in turn acts to depress the firm's profit possibilities. These ambiguities need analysis. Meanwhile a slight *negative* association between assets and profit rates has emerged

6. *Imports.* In recent decades, imports have probably provided the most important form of new competition in industrial markets in advanced economies. They may offer the best chances for applying Baumol et al.'s approach, even if the concepts may need adapting.

Import competition is defined by supply functions possessing continuous rather than quantum properties. Therefore imports are probably best analyzed in terms of internal market conditions, not of ultra-free entry.[28] Imports are difficult to define as a matter of entry. The outside producer does not fully *enter* the market, even though its products do. No new capacity is created in the market, and "exit" means merely a decline in the amounts shipped in, not a genuine closure. Because import entry reflects basic industrywide shifts in demand and supply (or in the market's edges), an analysis of strategic choices is largely irrelevant to it. In fact, the conventional entry analysis also adds little to research on the comparative costs and demand for imports. Still, this is probably the main research frontier for Baumol et al.

7. *Airlines.* I turn finally to the most promising single example of ultra-free entry; the airline industry. Under deregulation during 1975–84, the industry has displayed rising competition and more flexible pricing.

But this does not prove the validity or generality of ultra-free entry theory. One reason is that markets are not well defined, and so the roles of barriers, entry and exit are unclear. The *industry* covers air traffic in the United States, along hundreds of city-pair routes. At one extreme, one can define that national industry as the relevant *market*. But, at the other extreme, one might instead define *each city-pair route* as a relevant market. That is what Baumol et al. do, saying that each airline's addition of a route is entry into a new market. On this assumption, ultra-free entry may (nearly) exist: established airlines all act as potential entrants into each others' routes.

Yet many or most of these hundreds of city-pair routes are not relevant markets, by the standard criteria of substitutability. Many are paralleled closely by alternative routes which are close substitutes; many are merely intermediate stops in longer routes. Moreover, most routes are served as joint products; each flight on each route is only a small element in the airlines' total operations. Therefore most route changes are merely parts of multipoint competitive strategy in related market segments, *not* simple entry. Entry into the *industry* by founding a new airline is a much larger and slower task than merely adding routes. That form of entry has been far from total or nonresponded. Even simple route-addition "entry" has not been ultra free. Existing carriers have not been displaced at a stroke, and they have usually responded, often effectively. It takes time to establish ground facilities and build up patronage, usually by some kind of interaction involving prices and services. As market shares are gained and lost, the process accords well with mainstream analysis.

Moreover, few (if any) episodes of brief, profitable entry have been documented, as distinct from general impressions. Nor has systematic research established that ultra-free entry has effaced the role of internal structure. There is a significant relation between price-cost ratios and concentration in city-pair routes (see David Graham, Daniel Kaplan, and David Sibley, 1983).[29] High

---

in numerous partial regressions, while market share, concentration and certain entry-barrier variables have positive slopes. This suggests that the sunk cost element may have little net effect on price in the presence of other elements.

[28]For example, if imports cut General Motors' share from 58 to 44 percent, that is the key effect on market power.

[29]The authors note that their results are "inconsistent with the simplest version of the contestability hypothesis" (p. 137). Yet in Bailey et al., the tests are modified and a final chapter draws conclusions highly favorable to ultra-free entry. The first interpretation appears to be the more careful and reliable. Theodore Keeler (1981) offers further evidence favorable to internal conditions. Still other evidence of sharp fare differentials on monopolized routes (for example, of US Air) indicates that the cross-section regressions understate the true price impacts of market shares, despite legally free entry.

individual market shares are also associated with high price-cost ratios. Though ultra-free entry might enforce uniform price-cost ratios across all submarkets, in fact market power affects those ratios.[30]

Entry and exit have had some influence on the changes of market shares, but the main effects appear to arise from interactions among the existing airlines. As research on airlines has deepened, its support for ultra-free entry theory has grown weaker and now seems scarcely to exist. Airlines competition can be explained well by established concepts of market structure and entry.

IV. Summary

Industrial organization is primarily about internal market conditions. Baumol et al. have offered ultra-free entry as a new general system in place of the core concepts and evidence of the field. Instead their analysis only treats a specialized, extreme set of conditions, which are probably found in no real markets which have significant internal market power. Little has been added to the preexisting entry and exit analysis.

Thus no general basis has been provided, but rather an extended *gedankenexperiment*. Moreover the ultra-free entry model rests on assumptions which are contradictory (total entry and nonresponse) and which reverse reality. The more novel insights, such as entry-nullified monopoly, appear to be the least acceptable, on both logical and empirical grounds. Seen separately from the impressive analysis of multiproduct costs and prices, the ultra-free entry ideas are in urgent need of scientific support. Ultra-free entry may have value in the microeconomic theory curriculum, but it offers little so far to industrial organization research and teaching.

Several steps for research would be natural. Ultra-free entry analysis needs a thorough, critical evaluation, testing the assumptions, and defining the elements (such as barriers and entry) more fully and measuring them. Next one needs estimates of the relationships, so as to show the elements' relative importance and to test ultra-free entry's robustness. Thorough case studies are needed, and the normative analysis of ultra-free entry needs to be extended to include the other important performance elements, particularly efficiency in realistic settings, innovation, and equity.

In short, ultra-free entry faces a long, uphill research path, which may take much time and seems unlikely to rehabilitate the model. In the meantime, wise public policy choices will remain based on the accumulation of past research, with its focus on actual competition, as possibly modified by entry. Baumol et al.'s advice to avoid unnecessary entry barriers is sound, but it was already widely accepted. The "new" analysis gives no persuasive reason to shift attention away from competition within the market.

[30]These conclusions hold if the measures of cost are valid. It has been said, instead, that unit costs cannot be measured because of the overhead costs in operations spanning many routes. If that is so, then *no* tests of price-cost ratios can be made, either to prove or disprove entry's role. In short, airline experience either conflicts with Baumol et al.'s interpretation, or cannot be used to support it.

REFERENCES

Bailey, Elizabeth E. and Baumol, William J., "Deregulation and the Theory of Contestable Markets," manuscript, December 1982.

______, **Graham, David R. and Kaplan, Daniel P.,** *Deregulating the Airlines: An Economic Analysis*, Washington: Civil Aeronautics Board, May 1983.

Bain, Joe S., *Barriers to New Competition*, Cambridge: Harvard University Press, 1956.

______, *Industrial Organization*, rev. ed., New York: Wiley & Sons, 1968.

Baumol, William J., "Contestable Markets: An Uprising in the Theory of Industry Structure," *American Economic Review*, March 1982, *72*, 1–15.

______, **Bailey, Elizabeth E. and Willig, Robert D.,** "Weak Invisible Hand Theorems on the Sustainability of Prices in a Multiproduct National Monopoly," *American Economic Review*, June 1977, *67*, 360–65.

______, **Panzar, John C. and Willig, Robert D.,** *Contestable Markets and The Theory of Industry Structure*, San Diego: Harcourt

Brace Jovanovich, 1982.

______, ______, and ______, "Contestable Markets: An Uprising in the Theory of Industry Structure: Reply" *American Economic Review*, June 1983, *73*, 491–96.

Blair, John M., *Industrial Concentration*, New York: Harcourt Brace Jovanovich, 1972.

Bork, Robert H., *The Antitrust Paradox*, New York: Basic Books, 1978.

Brock, William A., "Contestable Markets and the Theory of Industry Structure: A Review Article," *Journal of Political Economy*, December 1983, *91*, 1055–66.

Caves, Richard E. and Porter, Michael E., "Barriers to Exit," in Robert T. Masson and P. David Qualls, eds., *Essays in Industrial Organization in Honor of Joe S. Bain*, Cambridge: Ballinger, 1976.

______ and ______, "From Entry Barriers to Mobility Barriers," *Quarterly Journal of Economics*, May 1977, *91*, 241–61.

Chamberlin, Edward H., *The Theory of Monopolistic Competition*, 8th ed., Cambridge: Harvard University Press, 1962.

Clark, John Bates, *Essentials of Economic Theory*, New York, Macmillan, 1907.

Clark, John M., *Studies in the Economics of Overhead Costs*, Chicago: University of Chicago Press, 1923.

Comanor, William S. and Wilson, Thomas A., "The Effects of Advertising on Competition: A Survey," *Journal of Economic Literature*, June 1979, *17*, 453–76.

Fellner, William J., *Competition Among the Few*, New York: Knopf, 1949.

Graham, David R., Kaplan, Daniel P. and Sibley, David S., "Efficiency and Competition in the Airline Industry," *Bell Journal of Economics*, Spring 1983, *14*, 118–38.

Heywood, John S., "Empirical Studies of Entry Barriers," manuscript, University of Michigan, October 1982.

Kahn, Alfred E., *The Economics of Regulation*, 2 vols., New York: Wiley & Sons, 1971.

Keeler, Theodore E., "The Revolution in Airline Regulation," in Leonard W. Weiss and Michael W. Klass, eds., *Case Studies in Regulation*, Boston: Little, Brown, 1981.

Kessides, Ioannes, "Toward a Testable Model of Entry: A Study of the U.S. Manufacturing Industries," manuscript, Princeton University, 1982.

Kwoka, John E., "The Effect of Market Share Distribution on Industry Performance," *Review of Economics and Statistics*, February 1979, *61*, 101–09.

Landes, William S. and Posner, Richard A., "Market Power in Antitrust Cases," *Harvard Law Review*, March 1981, *94*, 937–96.

Marshall, Alfred, *Principles of Economics*, 8th ed., London: Macmillan, 1920.

Martin, Stephen, *Market, Firm, and Economic Performance*, Monograph Series in Economics and Finance, New York University, 1983.

Neumann, John von and Morgenstern, Oskar, *Theory of Games and Economic Behavior*, Princeton: Princeton University Press, 1944.

Peltzman, Sam, "The Gains and Losses from Industrial Concentration," *Journal of Law and Economics*, October 1977, *20*, 229–63.

Posner, Richard A., "The Chicago School of Economic Analysis," *University of Pennsylvania Law Review*, April 1979, *127*, 925–48.

Ravenscraft, David J., "Structure-Profit Relationships at the Line of Business and Industry Level," *Review of Economics and Statistics*, February 1983, *65*, 22–31.

Robinson, Joan, *The Economics of Imperfect Competition*, London: Macmillan, 1933.

Scherer, F. M., *Industrial Market Structure and Economic Performance*, 2d. ed., Boston: Houghton Mifflin, 1980.

______, Open review letter on "Industry Structure Analysis and Public Policy," Northwestern University, June 15, 1982.

Schmalansee, Richard, "Another Look at Market Power," *Harvard Law Review*, June 1982, *95*, 1789–816.

Schwartz, Marius and Reynolds, Robert J., "Contestable Markets: An Uprising in the Theory of Industry Structure: Comment," *American Economic Review*, June 1983, *73*, 488–90.

Shepherd, William G., *Market Power and Economic Welfare*, New York: Random House, 1970.

______, "The Elements of Market Structure," *Review of Economics and Statistics*, February 1972, *54*, 25–37.

______, *The Treatment of Market Power*, New York: Columbia University Press,

1975.

______, *The Economics of Industrial Organization*, Englewood Cliffs: Prentice-Hall, 1979.

______, (1982a) "Causes of Increased Competition in the U.S. Economy, 1939–1980," *Review of Economics and Statistics*, November 1982, *64*, 613–26.

______, (1982b) "Competition and Sustainability," in Thomas G. Gies and Werner Sichel, eds., *Deregulation: Appraisal Before the Fact*, Ann Arbor: University of Michigan Business School, 1982.

______, (1982c) "Monopoly Profits and Economies of Scale," in John Craven, ed., *Industrial Organization, Antitrust, and Public Policy*, Boston: Kluwer Nijhof, 1982.

______, "Concepts of Competition and Efficient Policy in the Telecommunications Sector," in Eli Noam, ed., *Telecommunications Regulation Today and Tomorrow*, New York: Harcourt Brace Jovanovich, 1983.

Spence, A. Michael, "Entry, Capacity, Investment and Oligopolistic Pricing," *Bell Journal of Economics*, Autumn 1977, *8*, 534–44.

______, "*Contestable Markets and the Theory of Industry Structure*: A Review Article," *Journal of Economic Literature*, September 1983, *21*, 981–90.

Stigler, George J., *The Organization of Industry*, Homewood: Richard Irwin, 1968.

Weiss, Leonard W., "Optimal Plant Size and the Extent of Suboptimal Capacity," in Robert T. Masson and P. David Qualls, eds., *Essays on Industrial Organization: In Honor of Joe S. Bain*, Cambridge: Ballinger, 1976.

Weitzman, Martin L., "Contestable Markets: An Uprising in the Theory of Industry Structure: Comment," *American Economic Review*, June 1983, *73*, 486–87.

Williamson, Oliver E., *Markets and Hierarchies*, New York: Free Press, 1975.

Unattainability of Integrability and Definiteness Conditions in the General Case of Demand for Money and Goods

By Paul A. Samuelson and Ryuzo Sato*

Two fallacies have long permeated classical monetary theories: one, that demand for money is not subject to the analysis of marginal utility; two, that money is like coffee, music boxes, or any other good in its marginal utility analysis. One way to avoid these fallacies, not the only way and in the end not the best way, is to use the model of Samuelson (1947), which puts steady-state money and prices directly in the utility function along with steady-state goods like coffee and music boxes. This model does capture the difference between money and goods—the fact that money is not wanted for its own sake but only for the advantage of being able to make transactions—by insisting that increasing money and all prices in the same proportions will *ceteris paribus* leave the utility function unchanged. Thus the model does encapsulate the germ of truth in the Quantity Theory.[1]

Demand theory for ordinary goods has observable empirical implications. Pareto and Marshall began the investigation of these; Slutsky advanced the search by a quantum jump; and, finally in our modern times, *complete* observable implications of the corpus of demand theory have been derived, in that sense closing the theory for all times. No such closure has been achieved for demand theory inclusive of money (despite some imperfect attempts in the literature). The present paper suggests why such attempts are destined to fail in the *general* case. Mathematics has the useful function in science of helping establish what is true, what can be established. Mathematics can also perform the needed function of helping identify what cannot be established.

Does this analysis mean that a monetarist might be right in believing that the amount of M held could be independent of the interest rate r that provides the opportunity-cost hurdle that the desirability of holding a cash balance must overcome? No, our nihilistic presumption applies only to the demands for goods and leaves the usual Slutsky-Hicks presumptions concerning money intact. (More precisely, we establish that the demand for coffee might in a model involving money have positive income elasticity and still show Giffen-like positive price elasticity! Money, however, cannot show such perversity.)

This paper also considers a number of cases of "separable" demands in which *complete* integrability and definiteness conditions can be derived for observational testing, for refutation or nonrefutation. Also we generalize the model involving money and an interest rate, (M, r), to a model involving a vector of assets of differing degrees of liquidity, each element of which has a characteristic respective yield that defines the opportunity cost relevant to holding it: $(M_1,\dots, M_N; r_1,\dots, r_N)$.

Study of this money model yields, serendipitously as a byproduct, insight into the consequences of Veblen and Scitovsky phenomena in which tastes are themselves affected by relative prices.

*Departments of Economics, Massachusetts Institute of Technology, E-52, 383, Cambridge, MA 02139; and Brown University, Providence, RI 02912 and J. F. Kennedy School of Government, Harvard University, Cambridge, MA 02138, respectively. This work was in part supported by the National Science Foundation. We acknowledge useful comments on an earlier version by Robert Clower, John G. Riley, Jerome L. Stein, Fumitake Mimura, Don Patinkin, Peter W. Howitt, and Shunichi Tsutsui.

[1]In Section IX, we comment selectively on related work by C. E. V. Leser (1943), Don Patinkin (1956, 1965), Cliff Lloyd (1964), Per Meinich (1964), Richard Dusansky and P. J. Kalman (1972, 1974), Robert Clower and John Riley (1976), Clower and Peter Howitt (1978), Robert Pollak (1977), Howitt and Patinkin (1980), Dusansky (1980), and others. Also, in Section VII, we comment on the paper by R. L. Basmann, D. J. Molina and D. J. Slottje (1983).

I. The Money-Goods Model

Quantities of ordinary goods, coffee and such, we write as $(x_1,\dots,x_n)=\mathbf{X}$; their respective prices are $(p_1,\dots,p_n)=\mathbf{P}$. The steady-state stock of money is the zero-th good, M. To hold a unit of it in the steady state you must forego the positive yield on earning assets you might otherwise have held, r: thus, subtracted from your disposable steady-state income, Y, is not only p_1x_1 for your coffee expenditure but also rM for your expenditure incurred to hold the ("average") cash balance M.

You face the following maximum problem

$$(1)\quad \operatorname*{Max}_{M,\mathbf{X}} U[M,x_1,\dots,x_n;p_1,\dots,p_n]$$

$$= \operatorname*{Max}_{M,\mathbf{X}} U[M,\mathbf{X};\mathbf{P}] = \operatorname*{Max}_{M,\mathbf{X}} U[1,\mathbf{X};\mathbf{P}/M]$$

(2) subject to

$$rM+\sum_1^n p_jx_j = Y = rM+\mathbf{P}\cdot\mathbf{X}$$

$$(3)\quad = U[Yh^0(r;\mathbf{P},Y),h^1(r;\mathbf{P},Y),\dots,$$
$$h^n(r;\mathbf{P},Y);\mathbf{P}] = U[\mathbf{h}(r;\mathbf{P},Y);\mathbf{P}/Y]$$

$$(4)\quad = U^*[r;\mathbf{P},Y] = U^*[r;\mathbf{P}/Y,1],\qquad r>0,$$

where the demand functions $\mathbf{h}(r;\mathbf{P},Y)$ have the following contrasting homogeneities for goods and for money

$$(5a)\quad x_i = h^i(r;\mathbf{P},Y)\equiv h^i(r;\mathbf{P}/Y,1),\qquad (i=1,\dots,n);$$

$$(5b)\quad M/Y = h^0(r;\mathbf{P},Y)\equiv h^0(r;\mathbf{P}/Y,1)$$
$$= r^{-1}\left[1-\sum_1^n (P_j/Y)h^j(r;\mathbf{P}/Y,1)\right].$$

Note that coffee or any x_i continues in this money model to have the property of being homogeneous of degree zero in prices and income, as in conventional moneyless theory. By contrast, the demand for M is deduced to be homogeneous of degree *one* in prices and income (as is relevant to the ancient Quantity Theory of Copernicus, Locke, Hume, Thornton, Marshall, Ricardo, and Irving Fisher).[2]

To establish the validity of the many assertions in (3)–(5), we shall be making here some stronger simplifying assumptions than are needed in a truly general analysis. We can stay in the realm of calculus *interior equalities*, and out of the Kuhn-Tucker world of *boundary* extrema, by stipulating the following:

(6a) $U[M,\mathbf{X};\mathbf{P}]$ is strictly quasi concave in $[M,\mathbf{X}]$, having strictly convex indifference contours that are smoothly differentiable;

(6b) the various partial derivatives exist as continuous functions, and throughout we use numerical subscripts to a function to denote such derivatives

$$\partial U/\partial M = U_0[M,\mathbf{X};\mathbf{P}]$$
$$\equiv U_0[1,\mathbf{X};\mathbf{P}/M]M^{-1}>0$$

$$\partial U/\partial x_i = U_i[M,\mathbf{X};\mathbf{P}]$$
$$\equiv U_i[1,\mathbf{X};\mathbf{P}/M]>0,\qquad (i=1,\dots,n);$$

$$\partial U/\partial p_i = U_{n+i}[M,\mathbf{X};\mathbf{P}]$$
$$= U_{n+i}[1,\mathbf{X};\mathbf{P}/M]M^{-1},\qquad (n+i=n+1,\dots,2n);$$

$$\partial^2U/\partial M^2 = U_{00},\qquad \partial^2U/\partial x_i\,\partial x_j = U_{ij},$$
$$\partial^2U/\partial M\,\partial x_i = U_{0i},$$
$$\partial^2U/\partial p_i\,\partial x_j = U_{n+i,j},\dots;$$

$$(6c)\quad \mathbf{A}>0,\ U[\mathbf{A};\mathbf{P}]=U[\mathbf{B};\mathbf{P}]\quad \mathbf{B}>0.$$

[2]Many moneys used in history—gold is only one example—have a direct utility for their own sake as well as serving a transaction and store-of-wealth need. Marshall and early writers wouldn't expect such moneys to be "strictly neutral" in any run; when we speak of the simple Quantity Theory, we therefore refer to the polar case of fiat money or of a purely conventional material object that is of indifference for its own sake.

Relations (6) assure a *unique* interior solution to (1)'s maximum problem: strong convexity rules out ties and smoothness makes the h functions differentiable; (6c)'s "Inada" conditions rule out boundary maxima; etc.

Under our overstrong curvature conditions, the first-order *necessary* variational conditions are also *sufficient*. Avoiding mention of any nonobservable Lagrangean multiplier variables, we can write the maximizing conditions in a form that shows their invariance when $U[M,\mathbf{X},\mathbf{P}]$ is subjected to any monotone-increasing stretching, $u[M,\mathbf{X};\mathbf{P}] = f(U[M,\mathbf{X};\mathbf{P}])$ with $f'(U) > 0 \lesseqgtr f''(U)$:

$$\text{(7a)} \qquad \frac{U_i[M,\mathbf{X};\mathbf{P}]}{U_0[M,\mathbf{X};\mathbf{P}]} = \frac{p_i}{r}, \qquad (i=1,\dots,n);$$

$$\text{(7b)} \qquad rM + \sum_1^n p_j x_j = Y.$$

Using the homogeneity property of U in (1), and the implied (6) homogeneity properties of U's partials, we can rewrite (7) in a form that displays (5)'s assertions about the homogeneities of the $\mathbf{h}(r,\mathbf{P},Y)$ functions:

(8a)

$$\frac{U_i[M/Y,\mathbf{X};\mathbf{P}/Y]}{U_0[M/Y,\mathbf{X};\mathbf{P}/Y]} = \frac{p_i/Y}{r}, \quad (i=1,\dots,n);$$

$$\text{(8b)} \qquad r(M/Y) + \sum_1^n (p_j/Y)x_j = 1.$$

Under our assumptions, the $(n+1)$ equations of (8) can be solved for a unique solution for the $(n+1)$ real unknowns, $[M/Y, x_1,\dots,x_n]$ in terms of the $n+1$ prescribed real parameters $[r, p_1/Y,\dots,p_n/Y] = [r,\pi_1,\dots,\pi_n] = [r,\Pi]$.

II. Review of Conventional Moneyless Model

In the standard demand model, $U[M/Y, x_1,\dots,x_n; p_1/Y,\dots,p_n/Y]$ involves no p_j/Y parameters beyond the semicolon and $M/Y = x_0$ is an ordinary good with price $p_0/Y = r$; thus the partial derivatives $U_{n+i}[\ ]$ and $U_{j,n+i}[\ ]$ vanish in the moneyless model. It is well known for the moneyless model, under our strong simplifying assumptions, that its full empirical implications are summarized in the following $(n+1)$-by-$(n+1)$ negative semidefinite matrix:

Symmetry, for $i,j = 0,1,\dots,n$,

$$\text{(9)} \quad \mathbf{S} = \left[H_j^i(rY;\mathbf{P};Y) + H^j(rY;\mathbf{P};Y)H_{n+1}^i(rY;\mathbf{P};Y)\right] = \mathbf{S}^T$$

$$\text{where} \quad H_0^i = \partial H^i(p_0,\mathbf{P},Y)/\partial p_0,$$
$$H_{n+1}^i = \partial H^i/\partial Y, \text{etc.}$$

Homogeneity

$$\text{(10)} \qquad [p_0\mathbf{P}]\mathbf{S} = 0;$$

Negative semidefiniteness

$$\text{(11)} \quad \mathbf{Q}^T\mathbf{S}\mathbf{Q} < 0 \quad \text{for} \quad \mathbf{Q}^T \neq a[p_0\mathbf{P}]$$
$$\text{almost everywhere in } (r,\mathbf{P}/Y).$$

Duality theory can be used to deduce $\mathbf{S}$'s symmetry from the symmetry of U^*'s Hessian matrix. With no money in the model, Roy (1942) and others have shown

$$\text{(12)} \quad x_i = -\frac{U_i^*[p_0,p_1,\dots,p_n,Y]}{U_{n+1}^*[p_0,p_1,\dots,p_n,Y]}$$
$$= H^i[p_0,\dots,p_n,Y], \quad (i=0,1,\dots,n);$$

$$\text{(13)} \quad \mathbf{S} = \left[H_j^i + H^jH_{n+1}^i\right]$$
$$= \left(U_{n+1}^*\right)^{-3}\Big[-U_{ij}^*\left(U_{n+1}^*\right)^2$$
$$+ \left(U_i^*U_{n+1,j}^* + U_j^*U_{i,n+1}^*\right)\left(U_{n+1}^*\right)$$
$$- U_i^*U_j^*U_{n+1,n+1}^*\Big]$$
$$= \left[H_i^j + H^iH_{n+1}^j\right] = \mathbf{S}^T.$$

That $\mathbf{S}$ is singular follows from U^*'s homogeneity. That any principal minor of $\mathbf{S}$ is almost everywhere negative definite follows from U's strong quasi concav-

vity, with its implied similar property for $-U^*[p_0/Y,\ldots,p_n/Y,1]$.[3]

Unfortunately, once prices and money enter our model's utility function, it is conceptually impossible to unscramble contaminating effects of price on U_{n+i}/U_k. The simplicities of demand theory are lost.

IIII. Money's Contaminations

Duality theory rests on the basic envelope theorem of maximization theory, as in Samuelson (1947). To handle our many relations, we find it convenient to use the following Lagrangean expressions:

$$(14)\quad L=U[M,\mathbf{X};\mathbf{P}]+\lambda\left(Y-rM-\sum_1^n p_jx_j\right),$$

$$(15)\quad \partial L/\partial Y=\lambda=U^*_{n+1}[r;\mathbf{P},Y],$$

$$(16a)\quad \partial L/\partial r=-\lambda M=U_0^*[r;\mathbf{P},Y],$$

$$(16b)\quad M=-\frac{U_0^*[r;\mathbf{P},Y]}{U^*_{n+1}[r;\mathbf{P},Y]}=Yh^0(r;\mathbf{P}/Y,1),$$

$$(17a)\quad \partial L/\partial p_i=U_{n+i}[M,\mathbf{X};\mathbf{P}]-\lambda x_i=U_i^*[r;\mathbf{P},Y],$$

$$(17b)\quad x_i=-\frac{U_i^*[r;\mathbf{P},Y]}{U^*_{n+1}[r;\mathbf{P},Y]}+\frac{U_{n+i}[M,\mathbf{X};\mathbf{P}]}{\lambda}$$
$$=-\frac{U_i^*[r;\mathbf{P},Y]}{U^*_{n+1}[r;\mathbf{P},Y]}+Y\frac{U_{n+1}[M,\mathbf{X};\mathbf{P}]}{MU_0[M,\mathbf{X};\mathbf{P}]+\sum_1^n x_jU_j[M,\mathbf{X};\mathbf{P}]}$$
$$=h^i(r;\mathbf{P},Y)=h^i(r;\mathbf{P}/Y,1),\qquad (i=1,\ldots,n).$$

[3]See Sato (1981) for a more abstract version of the integrability conditions in terms of Lie groups.

Note that the breakdown between the two terms on (17b)'s right-hand side can generally not be made from observational data contained in the empirical $[h^i(r;\mathbf{P}/Y,1)]$ functions.

It follows that the following discrepancies generally will not vanish

$$(18)\quad [s_{ij}-s_{ji}]=\left[\left(h^i_j+h^jh^i_{n+1}\right)-\left(h^j_i+h^ih^j_{n+1}\right)\right].$$

Here $\partial h^i/\partial r=h^i_0(r;\mathbf{P},Y)$, etc. Moreover, these discrepancies involve the higher derivatives of the unidentifiable contamination terms, $U_{n+1}[M,\mathbf{X};\mathbf{P}]/U_k[M,\mathbf{X};\mathbf{P}]$.

In short, once money enters into the model in an essential way, demand theory not only loses its crown jewels (of testable well-behaved curvature and perhaps reciprocity relations), but worse than that, in a sense it loses its *raison d'etre* as a theory. The facts could be said to be left to tell their own story, except that our analysis has essentially demonstrated that the facts are left with no story to tell. All they can do is glare at us; all we can do is glare back and try to remember their multivariousnesses.

IV. Money and a Single Good

Here all is well behaved. We contemplate

$$(19)\quad \max_{M,x_1} U[M,x_1;p_1]=\max_{M/p_1,x_1} U[M/p_1,x_1;1]=\max_{x_0,x_1} U[x_0,x_1;1]$$

$$\text{subject to}\quad rM+p_1x_1=Y=\sum_0^1 p_jx_j$$

$$=U\left[H^0(p_0,p_1,Y),H^1(p_0,p_1,Y)\right]$$
$$=U\left[H^0(rp_1,p_1,Y),H^1(rp_1,p_1,Y)\right]$$
$$=U^*[r;p_1,Y]=U^*[r;p_1/Y,1].$$

Warning: Here $H^0=M/p_1$ and not M/Y as in (5b). It is apparent that, if we work

with $(x_0, p_0) = (M/p_1; rp_1)$, (19) becomes exactly the form of a standard moneyless system.

Hence, we can deduce the full empirical implications of such a standard system, which can be summarized by

$$\text{(20a)}\quad H^1(r; p_1, Y) = (Y/p_1) - rH^0(r; p_1, Y),$$

and almost everywhere,

$$\text{(20b)}\quad \frac{1}{p_1}\frac{\partial H^0(r; p_1, Y)}{\partial r} + H^0(r; p_1, Y)\frac{\partial H^0(r; p_1, Y)}{\partial Y} < 0.$$

Since money's "price," the interest rate r, does not enter explicitly into the utility function, M will always respond in the conventional function to a compensated *increase* in its price: that is, if we raise r, and at the same time raise Y by M times the change in r, utility will remain constant and there will always be a negative "substitution effect" in the Slutsky-Hicks fashion (no matter how **P** contaminates U!).

Equivalent to (20b) is

$$\text{(21)}\quad (\partial M/\partial r)_{\text{comp.}} = (\partial M/\partial r) + M(\partial M/\partial Y) < 0.$$

We shall see in Section VI that (21) is valid in the general case, no matter how large is n and without regard to special separability features.

The demand behavior for the good x_1 will, when $n = 1$, be just as well behaved as is the demand for money itself (for M or for x_0, as in (20b)). For $n = 1$, we can also write: almost everywhere

$$\text{(22a)}\quad 0 > \frac{\partial x_1}{\partial p_1} + x_1\frac{\partial x_1}{\partial Y} = h_1^1(r; p_1, Y) + h^1(r; p_1, Y)h_3^1(r; p_1, Y),$$

$$\text{(22b)}\quad 0 = (\partial U/\partial p_1)_{\text{comp.}}$$

The contaminating term(s) to the right of the semicolon in $U[M, \mathbf{X}; \mathbf{P}]$ can be got rid of for $n = 1$, when we use (19)'s trick of writing $U[M, x_1; p_1]$ as $U[M/p_1, x_1; 1] = U[x_0, x_1; 1]$.

We shall see that, although (20b) will hold for $n \geq 1$, alas, (22b) will not be assured for $n > 1$ when there are no Santa Claus separability properties.

V. Weak Separability

Complete integrability and Slutsky-curvature conditions can be derived in two cases; when the goods in the utility function are weakly separable, or when the prices in U are weakly separable. These respective cases are

$$\text{(23)}\quad U[M, \mathbf{X}; \mathbf{P}] \equiv U[M, g(\mathbf{X}); \mathbf{P}] = U[1, g(\mathbf{X}); \mathbf{P}/M]$$

$$\text{(24)}\quad U[M, \mathbf{X}; \mathbf{P}] \equiv U[M, \mathbf{X}; p(\mathbf{P})] = U[M/p(\mathbf{P}), \mathbf{X}; 1]$$

$$p(a\mathbf{P}) \equiv ap(\mathbf{P}), \quad \text{first-degree homogeneous.}$$

This second case, much discussed by Don Patinkin, Per Meinich, and others, easily reduces to the conventional moneyless model of our Section II. Assume we know the form of the $p(\mathbf{P})$ function and define a new set of left-hand variables:

$$\text{(25a)}\quad x_0 = M/p(\mathbf{P}), \quad \mathbf{X} = (x_1, \ldots, x_n), \quad p_0 = rp(\mathbf{P}), \quad \mathbf{P} = (p_1, \ldots, p_n).$$

Then

$$\text{(25b)}\quad p_0x_0 + \sum_1^n p_jx_j \equiv rM + \sum_1^n p_jx_j = Y,$$

$$\text{(25c)}\quad U[x_0, x_1, \ldots, x_n; 1] \equiv U[M, \mathbf{X}; p(\mathbf{P})] \equiv U[M/p(\mathbf{P}), \mathbf{X}; 1].$$

Looking at the left-hand sides of (25b) and (25c), we see only the components of a conventional moneyless demand model—in $n+1$ goods, consisting of n regular commodities

and one real-money magnitude. As in (12) and (13),

$$(26a)\quad x_i = H^i[p_0;\mathbf{P},Y],\ (i=0,1,\dots,n),$$

$$(26b)\quad \mathbf{S} = \left[H^i_j + H^j H^i_{n+1}\right] = \mathbf{S}^T,$$

$$(26c)\quad [p_0\mathbf{P}]\mathbf{S} = 0,$$

$$(26d)\quad Q^T\mathbf{S}Q < 0 \text{ almost everywhere, for } Q^T \neq a[p_0\mathbf{P}].$$

These tests of (26) can be applied to the observed $\mathbf{h}[r;\mathbf{P},Y]$ demand vector by utilizing the identities.

$$(27)\quad x_i = h^i[r;\mathbf{P},Y] \equiv H^i[rp(\mathbf{P});\mathbf{P},Y],\quad (i=1,\dots,n),$$

$$M = Yh^0[r;\mathbf{P},Y] \equiv p(\mathbf{P})H^0[rp(\mathbf{P});\mathbf{P},Y],$$

$$\left[H^i_0, H^i_j, H^i_{n+1}\right] = \left[h^i_0/p(\mathbf{P}), h^i_j - rh^i_0/p(\mathbf{P}) \times \{\partial p(\mathbf{P})/\partial p_j\}, h^i_{n+1}\right].$$

Warning: if $p(\mathbf{P})$ is not known to us in advance—and why would it be?—and why even be known to exist?—the (27) observations do not seem to be sufficient to "identify" the form or even existence of the $p(\mathbf{P})$ function and the (26) tests cannot be performed!

More satisfactory is the complete resolution of testable consequences of (23)'s weakly separable-quantities case. Here we see how to perform a two-stage maximization procedure.

1) Recognize first that the full optimum is not achieved unless the amount spent on real goods along, $\hat{Y} = Y - rM$, buys the consumer the highest level of $g(\mathbf{X})$. So, first solve the moneyless conventional problem (in n, not $n+1$, variables):

$$(28)\quad \underset{\mathbf{X}}{\mathrm{Max}}\, g(\mathbf{X}) \text{ subject to } \mathbf{P}\cdot\mathbf{X} = \hat{Y} = g(\mathbf{H}\{\mathbf{P},\hat{Y}\}).$$

To remind ourselves that the $\mathbf{H}\{\mathbf{P},\hat{Y}\}$ vector involves n and not $n+1$ elements, we use curly rather than square brackets. From Section II's moneyless theory we know that

$$(29)\quad \mathbf{S} = \left[H^i_j + H^j H^i_{n+1}\right] = \mathbf{S}^T,$$

$$\mathbf{PS} = 0,\ \mathbf{Q}^T\mathbf{SQ} < 0 \text{ for } \mathbf{Q}^T \neq a\mathbf{P} \text{ almost everywhere.}$$

2) In the second stage of the maximization procedure, with $(r,\mathbf{P},Y)$ regarded as fixed, we optimize how Y should be divided between M and $\hat{Y}$, recognizing how any $\hat{Y}$ will itself be optimally allocated among real commodities, and recognizing finally that the realization of both stages of the maximization must succeed in solving our full-optimality problem, so that

$$\underset{M,\mathbf{X}}{\mathrm{Max}}\, U[1, g(\mathbf{X});\mathbf{P}/M] \quad \text{subject to } rM + \mathbf{P}\cdot\mathbf{X} = Y$$

$$(30a)\quad = U[1, g(\mathbf{h}\{r;\mathbf{P},Y\});\mathbf{P}/M(r;\mathbf{P},Y)]$$

$$= \underset{M,\hat{Y}}{\mathrm{Max}}\, U[1, g(\mathbf{H}\{P,\hat{Y}\});\mathbf{P}/M] \quad \text{subject to } rM + \hat{Y} = Y,$$

$$(30b)\quad = U[1, g(\mathbf{H}\{\mathbf{P}, Y - rM(r,\mathbf{P},Y)\}); \mathbf{P}/M(r;\mathbf{P},Y)].$$

The reader who carefully studies the identity of (30a) to (30b) will be able to verify the two-stage-maximization identities

$$(31a)\quad x_i = h^i\{r;\mathbf{P},Y\} \equiv H^i\{\mathbf{P}; Y - rM(r;\mathbf{P},Y)\}.$$

If he or she differentiates both sides of these identities partially with respect to p_j, and then with respect to compensated changes in r, the reader will be able to solve explicitly

for the following:[4]

(31b) $$H^i_{n+1} = -r^{-1}\left(h^i_0 + Mh^i_{n+1}\right) / (M_r + MM_Y), \quad (i=1,\dots,n),$$

where $M_r + MM_Y < 0$, $r > 0$,

(32a) $$H^i_j = h^i_j - M_j\left(h^i_0 + Mh^i_{n+1}\right) / (M_r + MM_Y), \quad (i,j=1,\dots,n),$$

(32b) $$\mathbf{S} = \left[H^i_j + H^j H^i_{n+1}\right] = \left[h^i_j + h^j h^i_{n+1} - \left(M_j + h^j M_Y\right)\left(h^i_0 + Mh^i_{n+1}\right) / (M_r + MM_Y)\right] = \mathbf{S}^T,$$

negative semidefinite almost everywhere.

With some reflection the reader would note that (32b) is a modification of the simple Slutsky expression. This expression now includes the following: a compensated price change on money $(M_j + h^j M_Y)$, a compensated interest rate change on money $(M_r + MM_Y)$, and a compensated interest rate change on the ith good $(h^i_0 + Mh^i_{n+1})$.

These, at last, are $(n-1)(n-2)/2$ independent reciprocity conditions that the observable $\mathbf{h}(r; \mathbf{P}, Y)$ goods' demand functions must satisfy when $g(\mathbf{X})$ exists. Moreover, unlike $p(\mathbf{P})$, if the $g(\mathbf{X})$ function exists, that fact can be ascertained from the demand data; and those data provide us with the integrable differential forms that do "identify" the exact form of $g(\mathbf{X})$—of course, up to an arbitrary stretching of g, and with unambiguous marginal rate of substitution functions $g_j(\mathbf{X})/g_1(\mathbf{X})$ identifiable from p_j/p_1 data.

Worth mentioning is the following beautifully simple case that is weakly separable in both prices and quantities, and with $[p(\mathbf{P}), g(\mathbf{X})]$ functions that are first-degree homogeneous and are Shephard (1953) duals to each other. This case has the pleasing feature that it makes sense in the homothetic case for the person to possess a single price-index function which validly measures *at all income levels* the economists' notion of a true cost-of-living index. On the other hand, a painful price has to be paid for this logical consistency in that all experience tells us that real people do not have homothetic preferences.

Now we have

(33) $$U[M, \mathbf{X}; P] = U[M, g(\mathbf{X}); p(\mathbf{P})] = U[M/p(\mathbf{P}), g(\mathbf{X}); 1],$$

where

(34a) $$\operatorname*{Min}_{\mathbf{X}} \mathbf{P}\cdot\mathbf{X}/g(\mathbf{X}) = p(\mathbf{P})\text{: concave}$$
$$= a^{-1}p(a\mathbf{P})\text{: first-degree homogeneous}$$

(34b) $$\operatorname*{Min}_{\mathbf{P}} \mathbf{P}\cdot\mathbf{X}/p(\mathbf{P}) = g(\mathbf{X})\text{: concave}$$
$$= a^{-1}g(a\mathbf{X})\text{: first-degree homogeneous.}$$

From duality theory

(35a) $$x_i/g(\mathbf{X}) = [\partial \ln p(\mathbf{P})]/\partial p_i,$$

(35b) $$p_i/p(\mathbf{P}) = [\partial \ln g(\mathbf{X})]/\partial x_i.$$

Clearly, we can treat (33) as a two-good moneyless system

$$\operatorname*{Max}_{x_0, x_1} U[x_0, x_1; 1] \quad \text{subject to} \sum_0^1 p_j x_j = Y$$
$$= U^*[p_0, p_1, Y] = U^*[p_0/Y, p_1/Y, 1]$$

(36a) $$= U[H^0(p_0, p_1, Y), H^1(p_0, p_1, Y); 1]$$

(36b) $$\equiv \operatorname*{Max}_{M, g} U[M/p(\mathbf{P}), g; 1]$$

subject to $\quad rM + p(\mathbf{P})g = Y$

(36c) $$= U[M\{r; p(\mathbf{P}), Y\}/p(\mathbf{P}), g\{r; p(\mathbf{P}), Y\}; 1].$$

[4]Symbolically, each side of (31a) is subjected to $\partial/\partial p_j$ and to the compensated operation $(\partial/\partial r + M\partial/\partial Y)$. *Remark*: A compensated interest rate change on money $(M_r + MM_Y)$ is strictly negative for the general case as proved in Section VI (see 39e)).

In (36) $M\{r; p(\mathbf{P}), Y\}$ satisfies the single Slutsky-Hicks curvature condition

$$(37a)\qquad M_r + MM_Y < 0.$$

Also, from (35a), for $(i=1,\ldots,n)$,

$$(37b)\qquad x_i/g = h^i\{r;\mathbf{P},Y\}$$

$$/[Y - rM\{r; p(\mathbf{P}), Y\}]\, p(\mathbf{P})^{-1}$$

$$= \partial[\ln p(\mathbf{P})]/\partial p_i, \quad (i=1,\ldots,n),$$

with $[\partial^2\{\ln p(\mathbf{P})\}/\partial p_i\, \partial p_j]$ having to be symmetric and thereby implying $n(n-1)/2$ reciprocity condition on the $[\mathbf{X}/g, \mathbf{P}/Y]$ observable demand interrelations. Also, under our original strong quasi-concavity assumption, there will be (almost everywhere) the $n-1$ negative-semidefiniteness curvature conditions on the testable Hessian matrix of $\ln p(\mathbf{P})$.

VI. Well-Behaved Substitution Effect for Money

Invoking no separability properties for $U[M,\mathbf{X};\mathbf{P}]$ and letting there be any number of goods along with money, we can now show why a rise in the interest rate that is compensated by a rise in income just big enough to leave your utility intact must cause you to hold less money—a direct violation of extreme monetarism. The easiest approach is via the program of minimizing the expenditure needed to achieve a fixed level of utility:

$$(38a)\qquad \operatorname*{Min}_{M,\mathbf{X}} [rM + \mathbf{P}\cdot\mathbf{X}]$$

$$\text{subject to}\qquad U[M,\mathbf{X};\mathbf{P}] = \overline{U}$$

$$(38b)\qquad = E^*[r;\mathbf{P}|\overline{U}]$$

$$(38c)\qquad = rg^0(r;\mathbf{P}|\overline{U}) + \sum_1^n p_j g^j(r;\mathbf{P}|\overline{U}).$$

We wish to show that the $g^0_0(r;\mathbf{P}|U)$ demand function for money falls as the interest rate rises,

$$g_0^0 = \partial g^0(r;\mathbf{P}|\overline{U})/\partial r < 0 \text{ almost everywhere}$$

$$= Y\left[h_0^0(r;\mathbf{P},Y) + h^0(r;\mathbf{P},Y)\right.$$

$$\left.\times\{h^0_{n+1}(r;\mathbf{P},Y) + h^0(r;\mathbf{P},Y)\}\right]$$

$$= M_r + MM_Y < 0.$$

First, we prove this property for general finite changes in r alone, with $(p_1,\ldots,p_n)$ and U held constant at $(\overline{\mathbf{P}};\overline{U})$. Then for all $(M,\mathbf{X})$ that satisfy

$$(39a)\qquad U(M,\mathbf{X};\overline{\mathbf{P}}) = \overline{U}$$

to optimal g responses to $r=r^1$ and $r=r^2$, (M^1, X^1) and (M^2, X^2), satisfy

$$(39b)\qquad r^1M^1 + \overline{\mathbf{P}}\cdot\mathbf{X}^1 \le r^1M^2 + \overline{\mathbf{P}}\cdot\mathbf{X}^2,$$

$$(39c)\qquad r^2M^2 + \overline{\mathbf{P}}\cdot\mathbf{X}^2 \le r^2M^1 + \overline{\mathbf{P}}\cdot\mathbf{X}^1.$$

These inequalities, which will be strong inequalities if $r^1 \neq r^2$ and our earlier sections' regularity conditions are stipulated for U, follow from (38)'s defined minimum program.

Now if one subtracts respective sides of (39b) from those of (39c) and cancels, one derives

$$(39d)\qquad 0 \ge (r^2 - r^1)(M^2 - M^1)$$

$$= \Delta r \Delta M = (\Delta r)^2[\Delta M/\Delta r].$$

Again, the inequality will be a strong one for $r^2 \neq r^1$.

With M a smooth function of r, by letting $(\Delta r) \to 0$, we denote

$$(39e)\qquad \lim_{\Delta r\to 0}[\Delta M/\Delta r] = g_0^0(r;\mathbf{P}|\overline{U})$$

$$\equiv M_r + MM_Y < 0 \text{ almost everywhere.}$$

In conventional moneyless models, one can deduce for *any good* that compensated $\Delta p_j \Delta x_j$ is nonpositive. But only, so to speak, for $\Delta p_0 \Delta x_0$ can one validly arrive at this

conclusion once money enters the utility function along with its contaminating U_{n+i}/U_j effects. Thus, (39b) and (39c) would not be valid if disparate values of **P** were put into (39b) and (39c).

In summary, for a money model $g_j^j(r; \overline{P}|\overline{U})$ can be assured to be negative only for $j=0$.

VII. Theoretical Chaos when People Judge Quality by Price

Thorstein Veblen (1899) pointed out that people often want goods not so much because they are directly useful as because they are expensive. Tibor Scitovsky (1945) called attention to the fact that often we consumers judge the quality of goods by the prices charged for them: if Fords are cheap relative to Cadillacs and pork chops, we may regard that as a signal for the low worth of a Ford car.

Analytically, our indifference contours for (bread, Ford cars, pork chops) now depend on some relative prices, as on $p_{\text{Ford}}/p_{\text{pork}}$. Any indicator of utility, U, now becomes a function of price(s) as well as quantities:

$$(40)\quad U(\text{bread, Ford, pork}; p_{\text{Ford}}/p_{\text{pork}}) = U(x_0, x_1, x_2; p_1/p_2).$$

Although the context for this price contamination is different from that of our money models, the analytics look similar. Thus, recall a model involving two goods and money:

$$(41a)\quad U(M, x_1, x_2; p_1, p_2) \equiv U(M/p_2, x_1, x_2; p_1/p_2, 1),$$

$$(41b)\quad \equiv U(x_0, x_1, x_2; p_1/p_2, 1).$$

Comparison of (40) and (41b) shows them to be formally identical. When our previous analysis showed that theory can tell us virtually nothing useful about (41b)'s demand functions, $\mathbf{h}(p_0, p_1, Y)$, that demonstration warns us that theory is powerless to say interesting things about the Veblen-Scitovsky demand functions. As the elementary textbooks warn us, chic shops "may do better in moving a slow-selling item if they *raise* its price" (Samuelson, 1948, p. 474). We may now have $\partial x_1/\partial p_1 > 0$ even though $\partial x_1/\partial Y > 0$; and we aren't surprised if reciprocity conditions fail.

It will be instructive to consider two real goods, (x_1, x_2), whose marginal rates of substitution depend on the p_2/p_1 price ratio, $R[x_1, x_2; p_2/p_1]$:

$$(42a)\quad R[x_1, x_2; p_2/p_1] > 0 \quad \text{for } [x_1, x_2; p_2/p_1] > 0.$$

Almost everywhere in the positive (x_1, x_2) quadrant, no matter what fixed value p_2/p_1 takes, we posit

$$(42b)\quad 0 > R_2[x_1, x_2; p_2/p_1] - R[x_1, x_2; p_2/p_1]R_1[x_1, x_2; p_2/p_1] = \Delta[x_1, x_2; p_2/p_1].$$

Also

$$(42c)\quad \lim_{x_1 \to 0} R[x_1, x_2; p_2/p_1] = \infty, \qquad \lim_{x_2 \to 0} R[x_1, x_2; p_2/p_1] = 0.$$

Then, for each positive $(p_2/p_1, Y/p_1)$, there is a unique (x_1^*, x_2^*) demanded

$$(43)\quad x_1^* = h^1(1, p_2/p_1, Y/p_1) = h^1(p_1, p_2, Y),$$

$$x_2^* = h^2(1, p_2/p_1, Y/p_1) = h^2(p_1, p_2, Y).$$

These *direct* demand functions of (43) can be regarded as solutions of the *implicit* equations

$$(44a)\quad p_2/p_1 = R[x_1, x_2; p_2/p_1],$$

$$(44b)\quad Y/p_1 = x_1 + x_2 R[x_1, x_2; p_2/p_1].$$

But, of course, mere knowledge of the observable demand functions, of (43)'s

$h^i(p_1, p_2, Y)$ functions, does *not* give us knowledge of the $R[x_1, x_2; p_2/p_1]$.

(How could we get knowledge of $R[x_1, x_2; p_2/p_1]$ and of the indifference contours generated by any p_2/p_1 point? *Direct interrogation* of the consumer could give it in principle. For each p_2/p_1, he would compare for us any pair of (x_1, x_2) points. For p_2/p_1 given, he'd report

(a_1, a_2) better than (b_1, b_2), $\mathbf{A} \succ \mathbf{B}$

or (b_1, b_2) better than (a_1, a_2), $\mathbf{B} \succ \mathbf{A}$

or (a_1, a_2) and (b_1, b_2) indifferent, $\mathbf{A} \sim \mathbf{B}$.

Also, his $\mathbf{A} \succ \mathbf{B}$ and $\mathbf{A} \sim \mathbf{B}$ orderings would obey transitivity, etc.)

If the p_2/p_1 contamination of $R[\;]$ were "weak," so that $R_3[\;]$ is close enough to zero, (44a) could assuredly be solved for

$$(45) \qquad p_2/p_1 = f(x_1, x_2)$$

where $f(x_1, x_2)$ is a single-valued function. Also,

$$(46) \qquad f_2(x_1, x_2) - f(x_1, x_2)f_1(x_1, x_2) = \Delta(x_1, x_2)$$

will be assuredly negative wherever $R_3(x_1, x_2; p_2/p_1)$ is sufficiently near zero, since by the Implicit Function theorem

$$(47a) \qquad f(x_1, x_2) \equiv R[x_1, x_2; f(x_1, x_2)],$$

$$(47b) \qquad \Delta(x_1, x_2) \equiv \Delta[x_1, x_2; f(x_1, x_2)] / \{1 - R_3[x_1, x_2; f(x_1, x_2)]\}.$$

If (47b)'s numerator is required by (42b) to be negative, then where $R_3[x_1, x_2; f(x_1, x_2)]$ is sufficiently small, $\Delta(x_1, x_2)$ will be assuredly negative. For such *weak Veblen effects*, observed demand will still be well behaved.

However, let $R_3[x_1, x_2; p_2/p_1]$ become a sizable positive number. Then raising good 2's relative price will so much enhance good 2's relative marginal utility as to *reverse the curvature* of the reduced-form (pseudo!) indifference contours inferred by the naive observer of the (45) relation. Now $\Delta(x_1, x_2)$ has the opposite sign to $\Delta[x_1, x_2; f(x_1, x_2)]$ because (47b)'s denominator has become negative. No wonder that $\partial x_2/\partial p_2$ can be positive when $\partial x_2/\partial Y$ is positive: actually, the naive observer of (45) and (43) price-quantity interrelations will infer that he is observing a masochistic *minimizer* rather than a *homo economicus* who maximizes; each observed tangency of a budget line to a (pseudo) indifference contour will be a point of global minimum rather than maximum.

Qualification. There is no reason why R_3 shouldn't take on values that equal one at some places and that both exceed and fall short of one at other places. That means that no single-valued $f(x_1, x_2)$ function may exist. But, regardless of that, the *direct* demand functions of (43) are guaranteed to exist as single-valued smooth functions by our strong sufficiency conditions of (42) and of (6). The point is that they need not have conventional slope and curvature conditions.

With but two goods, there were no integrability conditions to be upset by Veblenesque effects. With three or more goods, the slightest Veblenesque effects will destroy the razor's-edge condition in which observable reciprocity effects necessarily obtain.

The case $n = 3$ will suffice. Consider the (not-directly-observable) *MRS* functions:

$$(48) \qquad p_2/p_1 = R^2[x_1, x_2; p_2/p_1],$$

$$p_3/p_1 = R^3[x_1, x_2, x_3; p_2/p_1],$$

$$Y/p_1 = x_1 + \sum_2^3 R^j[x_1, x_2, x_3; p_2/p_1]x_j.$$

Under our regularity conditions, (48) is compatible with

$$(49) \qquad x_j = h^j(p_1, p_2, p_3, Y) = h^j(1, p_2/p_1, p_3/p_1, Y/p_1), \qquad (j = 1,2,3).$$

For every positive p_2/p_1, the following two-by-two matrix is everywhere symmetric and

almost everywhere negative-definite:

$$(50a)\quad (a_{ij}) = \left(R^i_j - R^j R^i_1\right) = (a_{ji}).$$

Almost everywhere

$$(50b)\quad (q_1 \quad q_2)(a_{ij})\begin{pmatrix} q_1 \\ q_2 \end{pmatrix} < 0$$

$$\text{for } (q_1 \quad q_2) > 0.$$

If Veblenesque effects are "weak," so that $1-R^2_4$ is positive, the Implicit Function theorem assures us that (48) can be replaced by

$$(51)\quad p_j/p_1 = f^j(x_1, x_2, x_3) \quad (j=2,3)$$

$$Y/p_1 = x_1 + \sum_2^3 f^j(x_1, x_2, x_3)x_j.$$

However, it will *not* (generally) be the case that the reciprocity relation of (50a), $a_{12} = a_{21}$, will have the following valid counterpart

$$(52)\quad f^2_3(x_1, x_2, x_3) - f^3(x_1, x_2, x_3) f^2_1(x_1, x_2, x_3) = \Delta_{23}(x_1, x_2, x_3)$$

$$(53)\quad = \Delta_{32}(x_1, x_2, x_3).$$

Of course, $\Delta_{23} - \Delta_{32}$ will "almost vanish" if $R^2_4[x_1, x_2, x_3; f^2(x_1, x_2, x_3)]$ "almost vanishes." Also

$$(54)\quad \Delta_{22}(x_1, x_2, x_3) = f^2_2 - f^2 f^2_1$$

can be counted on to stay negative when R_4 remains sufficiently small. But as R_4 deviates more and more from zero, all bets are off concerning the signs of observed demand changes. Wherever $R^2_4[x_1, x_2, x_3; p_2/p_1]$ deviates from unity

$$(55)\quad \left(f^i_j - f^j f^i_1\right) = \left(R^i_j - R^j R^i_1\right) + \left(R^i_4\left[f^2_j - f^j f^i_1\right]\right).$$

Although the first matrix on the right (if it were observable) would have to be that of a symmetric matrix almost everywhere negative definite, the second (also unobservable) matrix generally spoils the conventional story. See, however, R. L. Basmann et al. for a special Fechner-Thurstone family of preferences, where the $f^j(\mathbf{X})$ functions are indistinguishable from the standard noncontamination case.

VIII. Multiple Money

Now replace a single money asset, M, with its opportunity cost of holding, r, by a whole vector of assets of different degrees of liquidity and usefulness for transaction purposes and steady-state holding: $(M_1, \ldots, M_k, \ldots, M_N) = \mathbf{M}$ with relevant interest-yield costs respectively of $(r_1, \ldots, r_k, \ldots, r_N) = \mathbf{R}$. Thus, in the steady state, I may hold some cash, M_1, with zero yield-return; its r_1 that measures what it costs me per dollar per annum in foregone income is a large positive amount. But, also, I hold a money market fund asset, M_2, that in the steady state adds to my utility; since, however, it yields me a high nominal return per annum—in 1981 a safe rate of as much as 18 percent—the f_2 that measures its foregone-income opportunity cost to me is lower than r_1 (possibly even being negative, depending on what convention we use for measuring opportunity costs). My demand problem is this: confronted by income Y, by prices of goods $\mathbf{P}$, and by vector of $\mathbf{R}$ costs for holding different moneys, I pick $(\mathbf{M},\mathbf{X})$ so as to maximize my steady-state utility.

We generalize our old one-money utility function of Samuelson (1947) in the following natural way:

$$(56)\quad U[M_1, \ldots, M_N, x_1, \ldots, x_n; P_1, \ldots, P_n] = U[\mathbf{M},\mathbf{X}; P] \equiv U[a\mathbf{M},\mathbf{X}; aP]$$

$$\partial U/\partial x_i = U_i[\mathbf{M},\mathbf{X};\mathbf{P}] > 0, \qquad (i=1,\ldots,n)$$
$$= U_i[a\mathbf{M},\mathbf{X}; a\mathbf{P}]$$

$$\partial U/\partial p_i = U_{n+i}[\mathbf{M},\mathbf{X};\mathbf{P}], \qquad (i=1,\ldots,n)$$
$$= U_{n+i}[a\mathbf{M},\mathbf{X}; a\mathbf{P}]a$$

$$\partial U/\partial M_k = U_{-k}[\mathbf{M},\mathbf{X};\mathbf{P}] > 0, \quad (k=1,\ldots,N)$$
$$= U_{-k}[a\mathbf{M},\mathbf{X}; a\mathbf{P}]a.$$

As before, U is to be strongly concave in $[\mathbf{M},\mathbf{X}]$; etc.

We write our $n+N$ demand functions for goods and for moneys as

$$(57)\quad x_i = h^i(r_1,\dots,r_N;p_1,\dots,p_n,Y), \quad (i=1,\dots,n)$$
$$= h^i(\mathbf{R};\mathbf{P},Y) = h^i(\mathbf{R};\mathbf{P}/Y,1)$$
$$M_k = Yh^{-k}(r_1,\dots,r_N;p_1,\dots,p_n,Y), \quad (k=1,\dots,N)$$
$$= Yh^{-k}(\mathbf{R};\mathbf{P},Y) = Yh^{-k}(\mathbf{R};\mathbf{P}/Y,1).$$

When there is only one money, $N=1$, our old $h^0(r;\mathbf{P}/Y,:1)$ becomes the same thing as our present notation's $h^{-1}(\mathbf{R},\mathbf{P}/Y,1)$.

These optimal amounts demanded, $(\mathbf{M}^*,\mathbf{X}^*)$, are the roots of the following variational equations:

$$(58\text{a})\quad U_i[\mathbf{M}/Y,\mathbf{X};\mathbf{P}/Y]b^{-1} = p_i/Y, \quad (i=1,\dots,n)$$

$$(58\text{b})\quad U_{-k}[\mathbf{M}/Y,\mathbf{X};\mathbf{P}/Y]b^{-1} = r_k, \quad (k=1,\dots,N)$$

$$b = \sum_{k=1}^{N} M_k U_{-k}[\mathbf{M},\mathbf{X};\mathbf{P}] + \sum_{j=1}^{n} x_j U_j[\mathbf{M},\mathbf{X};\mathbf{P}]$$
$$= \sum_{k=1}^{N}(M_k/Y)U_{-k}[\mathbf{M}/Y,\mathbf{X};\mathbf{P}/Y] + \sum_{j=1}^{n} x_j U_j[\mathbf{M}/Y,\mathbf{X};\mathbf{P}/Y].$$

The $(n+N)$ right-hand functions of (58) can be solved for our unique demand functions (57): these variational conditions are sufficient as well as necessary for solution to the following natural maximum problem.

$$(59\text{a})\quad \underset{\mathbf{M}\cdot\mathbf{X}}{\operatorname{Max}}\, U[\mathbf{M},\mathbf{X};\mathbf{P}] \text{ subject to } \mathbf{R}\cdot\mathbf{M}+\mathbf{P}\cdot\mathbf{X} = Y,$$
$$(59\text{b})\quad = U^*[\mathbf{R};\mathbf{P},Y] = U^*[\mathbf{R};\mathbf{P}/Y,1],$$
$$(59\text{c})\quad = U[\mathbf{h}(\mathbf{R};\mathbf{P}/Y,1);\mathbf{P}],$$

where $\mathbf{h}(\mathbf{R};\mathbf{P},Y) = [Yh^{-1}(\mathbf{R};\mathbf{P},Y),\dots, Yh^{-N}(\mathbf{R};\mathbf{P},Y), h^1(\mathbf{R};\mathbf{P},Y),\dots, h^n(\mathbf{R};\mathbf{P},Y)]$.

Students of rational macroeconomic policy will find confirmed in this neoclassical analysis their doubts concerning the validity of monetarists' prescriptions of simple rules. Should the Federal Reserve stabilize the trend-growth of M_1 or M_2 inclusive of money market fund assets? In the early 1980's, the crude empirical correlation between nominal *GNP* and M_2 exceeded the correlation between *GNP* and M_1. However, the relative sizes of these $\bar{R}^2$s would seem a poor guide to the kind of causality relevant for policy. The present model would explain a great 1979–81 increase in money market fund assets as a rational reaction to an increase in their nominal and real yields from one-digit to two-digit per annum rates; if that rise contributed toward a high $\bar{R}^2$ between *GNP* an M_2, this model would not expect that fact to increase confidence in M_2 as an important Granger-Sims causal factor of future *GNP* movements. In short, the present model is a model for understanding *variations* in various M_k *velocities* of circulation, not a model leading one to expect quasi constancy of such diverse velocities.

An important fruit of a theory is its suggestiveness concerning the likely pattern of comparative statics that will occur. Since the signs of $[\partial x_i/\partial p_i, \partial x_i/\partial Y]$ were seen to be nonidentifiable when the vector M consisted of a single element, we must, of course, expect that a multimoney model (with $N>1$) will suffer from the same deficiency.

Fortunately, the Slutsky-like matrix of M_k responses to changes in $[r_1,\dots,r_N]$ parame-

ters is as well behaved in the multimoney model as in the one-money model.[5] The following N-by-N matrix is symmetric and negative definite:

$$\text{(60a)}\quad \sigma=\left[\frac{\partial h^{-k}(\mathbf{R};\mathbf{P},Y)}{\partial r_q}+h^{-q}(\mathbf{R};\mathbf{P},Y)\frac{\partial h^{-k}(\mathbf{R};\mathbf{P},Y)}{\partial Y}\right]=\sigma^T,$$

$$\text{(60b)}\quad (z_1,\dots,z_N)\sigma(z_1,\dots,z_N)^T<0 \quad \text{for } (z_1,\dots,z_N)\neq 0.$$

No separate proof of this is needed. Since $\mathbf{P}$ is being held constant when elements of $\mathbf{R}$ change, and since no elements of $\mathbf{R}$ contaminate the U function, we may rely on classical moneyless demand analysis for results like (60) and for similar results on inverse demand functions in which $(M_1,\dots, M_N; p_1,\dots, p_n)$ are the independent variables. For such mixed-variables dualities, see Samuelson (1960).

The reader can explore special cases involving *weak separability*: as in the case of $N=1$, these special cases permit identifying testable reciprocity and curvature conditions.

What remains of the Quantity Theory when a vector of moneys replaces a single M? Its germ of truth still obtains. Every nominal price and wage, instead of being ultimately proportional to the quantity of externally prescribed M, now becomes a homogeneous (first-degree) function of the prescribed vector, $\mathbf{M}=(M_k)$. Only if one has an endogenous theory involving endogenous long-run invariances for the (M_k/M_1) ratios can a single scalar set the absolute scale for the system. Note that in a proper general equilibrium formulation there is no need for a Quantity Equation of Exchange with hypothetically constant velocity of circulation. Instead, the present neoclassical principles give an endogenous theory of how the velocity or velocities will vary in function of interest rates and other variables of the equilibrium.

[5]Robert Clower reminds us that what is called an increase in interest rates—some kind of a simultaneous increase in different kinds of interest rates—cannot be expected to have unambiguous sign responses on the part of each and every one of our many Ms. Agreed. What we can expect, and what the mathematical analysis confirms, is that a balanced (compensated) increase in any subset of the $[r_1,\dots, r_N]$ vector will, in the Hicksian fashion, have a definite negative effect on the Hicksian composite-commodity $\Sigma r_k M_k$, made up of the respective rs and Ms in question. No more can be expected to obtain. (All or some of the x_i goods flows might be in some degree inventoriable. So our vector of Ms might desirably be augmented by such inventory stocks. The present paper does not tackle the resulting problems in comparative statics.)

IX. False Starts

Here we shall merely mention some earlier studies that might be misinterpreted by hasty readers as having solved the problem shown here to be generally unsolvable.

One early model, that of C. E. V. Leser, seems to give a "generalized set of Slutsky equations" (1943, p. 130). However, the z_{rs} terms in them are precisely of the "contaminations" form we have noted in this paper, involving unobservable expressions that in our notation would be written as U_{n+i}/U_j. Moreover, we do not really understand this 1943 model. In our notation, it seems to be expressible as

$$\text{(61)}\quad \underset{x_1,\dots,x_n,L}{\text{Max}}\ U[x_1,\dots,x_n; L/p_1,\dots,L/p_n]$$

$$\text{subject to}\quad \sum_1^n p_j x_j + L = M.$$

In the steady state, if L does not equal M, we seem to deny the steadiness of the state; if L does equal M, $(x_1,\dots,x_n)$ can be a positive vector only if some other source of steady-state income is specified—wage income, interest income, or something. Modern dynamic analysis, of the kind that weds Miguel Sidrauski (1967) to Frank Ramsey (1928) would suggest a calculus-of-variations approach to a changing cash balance, $M(t)$. A typical program would involve

$$\text{(62)}\quad \underset{M(t),\mathbf{X}(t)}{\text{Max}}\ \sum_0^\infty U[M(t),\mathbf{X}(t);\mathbf{P}],$$

$$\text{subject to}\quad \sum_{j=1}^n p_j x_j(t)+rM(t) = Y(t)+M(t)-M(t-1).$$

Samuelson (1947, equations (117)) circumspectly refrained from trying to obtain

compensated demand coefficients. He was content to point out the well-behaved negative sign of the Slutsky term $(\partial M/\partial r)+M(\partial M/\partial Y)$.

Richard Dusansky and Peter Kalman (1972, 1974), citing writings of Samuelson, Lloyd, Meinich, Hadar, and Patinkin, derive Slutsky reciprocity relations. Despite certain words of exposition that seem to refer to a general $U[M,\mathbf{X};\mathbf{P}]$ function, it is clear that valid observationally testable results would crucially depend on the weak-separability cases—as for example their case of $U[M/p(\mathbf{P}),\mathbf{X};1]$ that was considered in our Section V. They also point out that when $U[M,\mathbf{X};\mathbf{P}]$ departs from homogeneity of degree zero in the variables $[M,\mathbf{P}]$ in such a way as to make the marginal rates of substitutions, U_{n-k}/U_n, also depart, then various "money illusions" are involved and the Quantity Theory's neutral money properties have to be qualified—a valid point too obvious to be itself controversial.

X. Final Reflections

The model that puts money and prices into the utility function in a homogeneous way has the merit of capturing some of the basic features of money. That doesn't mean it is an optimal model. It is not. Despite its helpfulness, we should be under no illusions on this point.

Money is useful to you in real life because it enables you to have a steadier and better adjusted time stream of consumptions: music boxes when and where you want to hear music; time profiles of coffee consumption with less variance because you don't run out of liquid cash to buy; better bargains available for ready cash; less uncertainty as to where your next meal is to come from, and fewer past disappointments concerning hoped for consumption goods. A really deep analysis, therefore, will put in its utility functions time profiles of real consumption goods: $[\mathbf{X}(t),\mathbf{X}(t+1),\dots,\mathbf{X}(t+T),\dots]$.

A larger average steady-state money balance,

$$(63)\qquad \overline{M}=T^{-1}\sum_{t=0}^{T}M(t),\qquad T\to\infty$$

will give a higher expected value to

$$(64)\qquad U[\dots,X(t),X(t+1),\dots].$$

But, as an ultimate *desideratum*, money is no more in the utility function than is penicillin or any other intermediate good. We must, therefore, regard our $U[M,\mathbf{X};\mathbf{P}]$ function as already the result of some behind-the-scenes optimizing and time-averaging. With this understanding and caveat, we can regard the present model as a useful scaffolding of analysis even if it is not itself at the deepest level.

Some other qualifications should be made explicit. This paper has analyzed effects of a change in the interest rate on $\mathbf{X}$ real-good consumptions or on M money holding while "income" is being held constant. Actually, a permanent rise in the interest rate might itself induce a rise in the *rentier's* steady-state income. (Or, for one whose sole asset is an acre of land with a fixed perpetual real annual payment, a rise in r will cause *no* change in steady-state Y, the induced drop in land's capitalized value just matching the rise in r and leaving annual spendable Y intact.) We make no pronouncement on these incidence questions. If you know the form of the $\mathbf{h}(r;\mathbf{P},Y)$ functions, then you are in a position to analyze any combination of $(r;\mathbf{P},Y)$ changes that are presented to you.

Ever since 1947, a number of writers have been bothered by the present steady-state analysis. "Why," they may ask, "does rM enter on the left-hand side of (2)'s budget equation, with M entering nowhere else?" And actually a number of writers have tried to follow the example of J. R. Hicks (1939) and of Don Patinkin (1956) and envisage a model involving successive weekly periods. At the beginning of the week, you have begun with $M(t)$ of money holding. Your "income" that you can spend in the week (paid, say, at the beginning of the week) is $Y(t)$. You must decide how much of goods, $\mathbf{X}(t)$, you'll buy to consume during the week; and you must decide how much $M(t+1)$ to end the week with. Provided prices are constant over time, $P(t)\equiv\mathbf{P}$, all you need is a utility function that orders the $[M(t+1),\mathbf{X}(t)]$ vector. Under stationary conditions your maximum problem becomes, for any

given t,

$$\text{(65a)}\quad \underset{M(t+1),\mathbf{X}(t)}{\text{Max}} U[M(t+1),\mathbf{X}(t);\mathbf{P}] = \underset{M(t+1),\mathbf{X}(t)}{\text{Max}} U[1,\mathbf{X}(t);\mathbf{P}/M(t+1)],$$

subject to

$$\text{(66)}\quad M(t+1)+\mathbf{P}\cdot\mathbf{X}(t)=M(t)+Y(t),$$

$$\text{(65b)}\quad =U^*[1;\mathbf{P},M(t)+Y(t)] = U^*[1;\mathbf{P}/\{M(t)+Y(t)\},1],$$

$$\text{(65c)}\quad =U[\{M(t)+Y(t)\} \times h^0(1;\mathbf{P}/\{M(t)+Y(t)\}, \mathbf{h}(1;\mathbf{P}/\{M(t)+Y(t)\}].$$

Our point is that the U in (65a) is *exactly* of the form analyzed in this paper; the *indirect utility* function in (65b) is *exactly* of the form analyzed here, except, of course, that the budget equation in (66) puts $M(t)+Y(t)$ where we had put Y and it puts our r equal to unity; and, finally, the optimizing demand functions in (65c) are *exactly* of the form we have already derived, once $[r;\mathbf{P},Y]$ are replaced as arguments by $[1;\mathbf{P},M(t)+Y(t)]$.[6]

Therefore, our present analysis handles both the Samuelson (1947) steady-state model and the Patinkin-Hicks week-by-week model. All of our negative results on identifiability apply to *both models*.

A further negative word needs to be said about the week-by-week model. In the form given it has defects that can be removed. And it also has intrinsic defects. First, writers like Cliff Lloyd (1964) and Dusansky and Kalman (1972; 1974) ought preferably to put an interest rate, r, somewhere into their analysis. Instead of taking $Y(t)$ income as given for the week, one would take the week as beginning with a stock of *total* assets $A(t)$. This can be "spent" on the week's real goods, $\mathbf{P}\cdot\mathbf{X}(t)$; or it can be held, all or in part, as an earning asset, $K(t)$, earning $rK(t)$ extra by the end of the week; or part of it can be held during the week as zero-yielding useful cash, $M(t)$, desired for its "convenience" in making $\mathbf{P}\cdot\mathbf{X}(t)$ transactions; or what's left over can be carried into next week's initial assets $A(t+1)$.

Now we can build up our "budget equation" from the identity:

$$\text{(67)}\quad A(t+1)=A(t)+[A(t)-\tilde{M}(t)]r + \tilde{Y}(t)-\mathbf{P}\cdot\mathbf{X}(t),$$

where $\tilde{Y}(t)$ is the period's noncapitalizable extraneous income, and $r[A(t)-\tilde{M}(t)]$ is the yield of earning assets. The budget relation becomes at the equilibrium $A(t)=A(t+1)$

$$\text{(68)}\quad r\tilde{M}(t)+\mathbf{P}\cdot\mathbf{X}(t)=\tilde{Y}(t)+A(t)r.$$

Then, subject to this,

$$\text{(69)}\quad \underset{\tilde{M}(t),\mathbf{X}(t)}{\text{Max}} U[1,\mathbf{X}(t);\mathbf{P}/\tilde{M}(t)]$$

will yield the observable demand functions:

$$\text{(70a)}\quad \tilde{M}=\{\tilde{Y}(t)+rA(t)\} \times h^0(r;\mathbf{P}/\{\tilde{Y}(t)+rA(t)\},1),$$

$$\text{(70b)}\quad x_i=h^i(r;\mathbf{P}/\{\tilde{Y}(t)+rA(t)\},1) \quad (i=1,\dots,n).$$

Remark: $K(t)$ could itself be put in the utility function along with M; your earning assets could also be one kind of "liquidity."

Note that the present dynamic formulation avoids the flaw involved in the conventional week-by-week model of (65a); in that model, determining terminal $M(t+1)$ essentially involves determining how much to *save* and yet it gives the reader the impression that its decisions have to do with prudent cash-balance management. In (65a), the dependence of U on $M(t+1)$ is not so much a matter that hinges on how near you are to a bank, how liquid your portfolio of stocks

[6] If a steady state is to exist uniquely for $Y(t)=Y$, one would have to equate the given $M(t)$ to the optimized $M(t+1)$ at M^*, the root of

$$M=(Y+M)h^0(1;\mathbf{P}/[Y+M],1).$$

and bonds is, or how unsynchronized and unpredictable are your outpayments and inpayments; the dependence of U on $M(t+1)$ there could merely reflect how much you weigh subsequent weeks' needs against that of the current week (will you be retired then? Do you have a Fisher-Böhm subjective time preference rate that is high or low?...).

Warning: Avoid the temptation to think of the week's money holding, $\tilde{M}(t)$, as being quantitatively the *same* thing as the $\Sigma p_j x_j(t)$ spent during the week. Twins might be equally well off when Twin A spends \$100 on $\mathbf{P}\cdot\mathbf{X}(t)$ and holds on the average \$10 in barren cash, while Twin B spends \$95 on $\mathbf{P}\cdot\mathbf{X}(t)$ and holds \$20 in barren cash. To let them be indifferent, Twin B's time profile of $\mathbf{X}(t)$ must be supposed to be a more "convenient" one than Twin A's profile of $\mathbf{X}(t)$. If the reader does not understand this distinction, he misses the point of the $U(1,\mathbf{X};\mathbf{P}/M)$ model.

Here we have analyzed the consumer as *price taker*. For a neoclassical closed model of monetary general equilibrium, see Samuelson (1968). For weak money "contaminations" presumably the usual existence theorems on existence of a competitive equilibrium will serve for a money economy with some appropriate modifications.[7]

[7]The following example serves as a warning that a system may have no positive-price equilibrium. Suppose all people are alike in maximizing an infinite lifetime Ramsey utility integral, with integrand

$$Ue^{-\bar{r}t} = [(p_1/p_2)\ln x_1 + \ln x_2 + M/p_2]e^{-\bar{r}t}.$$

They each have the same positive money and goods endowments initially, $(\bar{x}_1, \bar{x}_2, \overline{M})$. Then

$$p_1/p_2 = (\partial U/\partial x_1)/(\partial U/\partial x_2) = (p_1/p_2)(x_2/x_1)$$
$$r/p_2 = (\partial U/\partial M)/(\partial U/\partial x_2) = x_2/p_2.$$

If $\bar{x}_1 \neq \bar{x}_2$, there cannot be a positive-price equilibrium.

This case does not violate the general existence theorem of Lionel McKenzie (1955), for the reason that he permits an equilibrium to involve a zero price—as in the following when $\bar{x}_1 > \bar{x}_2$:

$$0 = (p_1/p_2)^* = (p_1/p_2)^*(x_2^*/x_1^*), \quad p_1^* = 0,$$
$$r^*/p_2^* = \bar{r}/\bar{p}_2 = x_2^*/M^*$$
$$x_2^* = \bar{x}_2, \quad M^* = \overline{M}, \quad x_1^* \leq \bar{x}_1, \quad p_2^* = \bar{r}\overline{M}/\bar{x}_2.$$

In concluding, we call attention to the open question whether, in the general nonseparable case, there are any testable consequences on the (5) observable demand functions of the (1)–(2) money-and-prices-in-the-utility-function model. Since there are such consequences for the case of money and a single good, and for the case of money and a weakly separable composite of goods, one will not give up hope that there exist such testable consequences until a definite *impossibility theorem* is proved or is shown to be false.

REFERENCES

Basmann, R. L., Molina, D. J. and Slottje, D. J., "Budget Constraint Prices as Preference Changing Parameters of Generalized Fechner-Thurston Direct Utility Functions," *American Economic Review*, June 1983, *73*, 411–13.

Clower, Robert W. and Howitt, Peter W., "The Transactions Theory of the Demand for Money: A Reconsideration," *Journal of Political Economy*, June 1978, *86*, 449–66.

______ **and Riley, John G.,** "The Foundations of Money Illusion in a Neoclassical Micro-Monetary Model: Comment," *American Economic Review*, March 1976, *66*, 184–85.

Dusansky, Richard, "Utility Function Transformations and Money Illusion: Reply and Further Results," *American Economic Review*, September 1980, *70*, 823–25.

______ **and Kalman, Peter J.,** "The Real Balance Effect and the Traditional Theory of Consumer Behavior: A Reconciliation," *Journal of Economic Theory*, December 1972, *5*, 336–47.

______ **and** ______, "The Foundations of Money Illusion in a Neoclassical Micro-Monetary Model," *American Economic Review*, March 1974, *64*, 115–22.

Hicks, J. R., *Value and Capital*, Oxford: Clarendon Press, 1939.

Howitt, Peter W. and Patinkin, Don, "Utility Function Transformations and Money Illusion: Comment," *American Economic Review*, September 1980, *70*, 819–22.

Leser, C. E. V., "The Consumer's Demand for Money," *Econometrica*, April 1943, *11*,

123–40.

Lloyd, Cliff, "The Real-Balance Effect and the Slutsky Equation," *Journal of Political Economy*, June 1964, *72*, 295–99.

McKenzie, Lionel, "Competitive Equilibrium with Dependent Consumer Preference," in *The Second Symposium on Linear Programming*, National Bureau of Standards and Department of the Air Force, Washington, 1955.

Meinich, Per, "Money Illusion and the Real Balance Effect," *Statsøknomisk Tideskrift*, 1964, *28*, 9–16, 20–21.

Patinkin, Don, *Money, Interest, and Prices*, 1st; 2nd eds., Evanston: Row, Peterson and Company, 1956; 1965.

Pollak, Robert A., "Price Dependent Preferences," *American Economic Review*, March 1977, *67*, 64–75.

Ramsey, Frank P., "A Mathematical Theory of Saving," *Economic Journal*, December 1928, *38*, 543–59.

Roy, R., *De l'utilité*, Paris: Herman, 1942.

Samuelson, Paul A., *Foundations of Economic Analysis*, Cambridge: Harvard University Press, 1947 (Enlarged edition, 1983).

______, *Economics*, 1st edition, New York: McGraw-Hill, 1948.

______, "Structure of a Minimum System," in Ralph W. Pfourts, ed., *Essays in Economics and Econometrics: A Volume in Honor of Harold Hotelling*, Chapel Hill: University of North Carolina Press, 1960, reproduced in J. E. Stiglitz, ed., Vol. 2, *The Collected Scientific Papers of Paul A. Samuelson*, Cambridge: MIT Press, 1972, ch. 44.

______, "What Classical and Neoclassical Monetary Theory Really Was," *Canadian Journal of Economics*, February 1968, *1*, 1–15, reproduced in J. E. Stiglitz, ed., Vol. 3, *The Collected Scientific Papers of Paul A. Samuelson*, Cambridge: MIT Press, 1972, ch. 176.

______ **and Swamy, S.,** "Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis," *American Economic Review*, September 1974, *64*, 566–93.

Sato, Ryuzo, *Theory of Technical Change and Economic Invariance: Application of Lie Groups*, New York: Academic Press, 1981.

Scitovsky, Tibor, "Some Consequences of the Habit of Judging Quality by Price," *Review of Economic Studies*, No. 2, 1945, *12*, 100–05.

Shephard, R. W., *Cost and Production Functions*, Princeton: Princeton University Press, 1953.

Sidrauski, Miguel, "Rational Choice and Pattern of Growth in a Monetary Economy," *American Economic Review Proceedings*, May 1967, *57*, 534–44.

Slutsky, E., "Sulla Teoria del Bilancio del Consumatore," *Giornale degli Economisti*, 1915, *51*, 1–26, translated as "On the Theory of the Budget of the Consumer," in G. L. Stigler and K. Boulding, eds., *Readings in Price Theory*, Homewood: Richard Irwin, 1952.

Veblen, Thorstein, *The Theory of the Leisure Class: A Study in the Evolution of Institution*, London; New York: Macmillan, 1899.

International Trade, Foreign Investment, and the Formation of the Entrepreneurial Class

By GENE M. GROSSMAN*

An important aspect of the development process in market-oriented economies is the formation of a "class of entrepreneurs"—that is, a group of individuals who are capable of organizing production and are willing to bear the risks associated with industrial activity. These agents serve an essential function in many less developed countries, because the markets and infrastructure necessary for the efficient and widespread allocation of risk across the population are often imperfect or completely absent. In cases where the supply of entrepreneurs remains limited, there may occur what Walt Rostow (1956) has called "secular stagnation," with little if any economic growth, and a tendency for overspecialization in the agricultural and traditional sectors. This point was argued most forcefully by W. Arthur Lewis:[1]

> Outside the sphere of agriculture, which can be conducted on a family size basis, economic growth is bound to be slow unless there is an adequate supply of entrepreneurs looking out for new ideas, and willing to take the risk of introducing them. Thus a private enterprise economy will be retarded if it has not enough business men, or if its business men are reluctant to take risks, whether because they cannot raise the capital, or because they are timid by nature, or because the differentials for risk-taking are inadequate. [1955, p. 182]

It has been further claimed by some (see, for example, Felipe Pazos, 1967, and Albert Hirschman, 1969) that openness to international competition, in the forms of free international trade, and especially inflows of direct foreign investment, serves to impede the development of the entrepreneurial class, and thus can be detrimental to the economy as a whole. Quoting from Hirschman,

> The opponents of free trade have often pointed out that for a variety of reasons it is imprudent and harmful for a country to become specialized along certain product lines in accordance with the dictates of comparative advantage. Whatever the merits of these critical arguments, they would certainly acquire overwhelming weight if the question arose whether a country should allow itself to become specialized not just along certain commodity lines, but along factor-of-production lines. Very few countries would ever consciously wish to specialize in unskilled labor, while foreigners with a comparative advantage in entrepreneurship, management, skilled labor and capital, took over these functions, replacing inferior "local talent." [p. 5]

An implication that has been drawn is that (at least) temporary restrictions to commodity trade and inward foreign investment can be justified on these grounds. Such intervention would not necessarily be inconsistent with the tenets of welfare economics as applied to trade policy, since the assumed starting point is one with an incomplete market structure, and thus falls under the rubric of "the theory of distortions" (see Jagdish Bhagwati, 1971).

Indeed, it is now known that in some situations of uncertainty, free trade may not be superior to autarky, and that tariffs may be welfare-improving even for a small economy (see, for example, David Newbery and

*Woodrow Wilson School, Princeton University, Princeton, NJ 08544. The first draft of this paper was completed during an enjoyable visit at the Institute for International Economics Studies, University of Stockholm. I gratefully acknowledge helpful discussions with Alan Deardorff, Avinash Dixit, Elhanan Helpman, and Henrik Horn, and the financial support of the National Science Foundation under grant SES-8207643.

[1]Similarly, Rostow argued that "...it is evident that the take-off requires the existence and the successful activity of some group in the society which accepts borrower's risk..." (p. 41). For further discussion, see Peter Kilby (1971).

Joseph Stiglitz, 1981; Jonathan Eaton and myself, 1981; and my forthcoming article). Trade may cause the inefficiency associated with an initially suboptimal allocation of risk across agents to be exacerbated, and thus generate a deleterious side effect that offsets the direct gains from specialization. In the present context, for example, free trade might be harmful if it *ceteris paribus* lowered the return to the entrepreneurial activity, and thereby caused fewer individuals to choose to become entrepreneurs, from an initial situation in which the supply of this factor was already suboptimally small. It seems, therefore, that a case might be made on this basis for sheltering (to some extent) domestic industrial enterprises in the less developed economies from exposure to the competitive forces of international trade and foreign investment until such a time as factor-supply conditions allow them to compete with the industrialized world on a more equal footing.

Compelling as these arguments may appear to be, it is nonetheless necessary to subject them to careful analytical scrutiny. The purpose of the present paper, then, is to explore in the context of a formal model some of the implications of "openness," for an economy in which domestic markets for risk sharing are absent and the supply of entrepreneurs is endogenous. Drawing on the work of S. M. Kanbur (1979, 1981), I develop in Section I a two-sector model of the open, less developed economy. There it is shown, in accordance with the verbal treatments cited above, that the free-trade equilibrium is characterized by an undersupply of entrepreneurs and an excessive degree of specialization in agricultural (or traditional) production, relative to a first-best social optimum.

In Section II, I compare the free-trade outcome with the equilibrium under complete autarky. First I investigate the effects of international trade on the welfare levels of the various "classes" in society, in the absence of any government redistributional policy. Then the question of potential gains from trade is addressed, with explicit consideration of whether trade creates an opportunity, via a feasible compensation scheme, for Pareto welfare improvement.

The efficacy of interventionist trade policy is examined in Section III. I show that protection from foreign competition can indeed be effective as a means of enlarging the pool of domestic entrepreneurs. However, the conclusions I draw regarding the welfare implications of such policies are considerably less sanguine.

Section IV deals with direct foreign investment, taken here to mean the establishment of foreign-owned enterprises in the less developed economy. I analyze the effects of an inflow of foreign firms on the size of the local entrepreneurial class, and on national income and welfare in the host country. The main findings of the paper are summarized in a concluding section.

I. The Model

I wish to endogenize the supply of entrepreneurs in an open-economy, general equilibrium setting. For this purpose, I borrow from Kanbur (1979, 1981), who has developed a model of occupational choice in which individuals face an *ex ante* decision whether to join a class of risk bearers or work instead for a safe (i.e., certain) wage.[2] This approach to entrepreneurship can be embedded in a simple, familiar specification of intersectoral resource allocation and international trade, so as to provide a tractable framework for addressing formally the issues raised in the introduction.

Consider then a small economy comprising an agricultural or "traditional" sector, and an industrial or "modern" sector. The output in agriculture is denoted by x, and this good is chosen as numeraire. Industrial output is z, with an (exogenous) relative price of p on world markets.

Production in agriculture requires the input of labor L_x and land T, according to $x = g(L_x, T)$. Labor earns a wage w (in units of the x good). The return to the fixed supply of land accrues to a predetermined group of rentiers, whom I shall refer to as the *landlord class*. The input of labor is assumed to be

[2] Richard Kihlstrom and Jean-Jacques Laffont (1979) present a model of entrepreneurship and risk taking which is quite similar to that of Kanbur.

governed by rent maximization by landlords, and satisfies

$$(1) \qquad g_L(L_x, T) = w.$$

Production in the industrial sector is organized and managed by a class of *entrepreneurs*. This undertaking involves risk, that may stem either from uncertainty regarding the ability of the entrepreneur as a manager or from some inherent aspect of the production process. The output of the ith enterprise is $f(l_z^i, \alpha^i)$, where α^i is a random variable and l_z^i is the labor hired by this entrepreneur (for simplicity, I assume that each firm or enterprise has exactly one entrepreneur). At the time that production takes place, labor is assumed to be perfectly mobile between firms and sectors, so that all labor in the industrial sector must be paid the (safe) competitive wage, w.[3] The entrepreneur, then, must bear the production (and hence profit) risk, and it is assumed that there do not exist markets on which he can purchase insurance.

Let the number of entrepreneurs in the economy be N, and suppose that the individual uncertainties are stochastically independent, but governed by the same density function, $\Phi(\alpha)$.[4] Each entrepreneur seeks to maximize his profits given the realization of the relevant random variable; thus, employment in the ith production unit is determined by

$$(2) \qquad pf(l_z^i, \alpha^i) = w.$$

It remains only to specify the occupational choice decision. The total population excluding landlords is L. Each of these individuals may opt to become a *laborer*, in which case a nonstochastic income of w is ensured, or else may join the entrepreneurial class, and bear the associated profit risk. There is a fixed cost, k, associated with entrepreneurship. This can be thought of as a set-up cost which must be borne *ex ante* (i.e., prior to the time when the prospective entrepreneur learns the value of α^i that would be applicable for him). The "loan" is repaid after production takes place, so that if the entrepreneur remains in business during the output stage, his net output $f(l_z^i, \alpha^i)$ equals $g(l_z^i, \alpha^i) - k$, where $g(\cdot)$ represents gross output. The entrepreneur alternatively might choose to abandon ship after he discovered that he lacked managerial ability or that his luck was bad. In this case his income would be $w - k$. However, I assume that k is sufficiently large relative to the range of output variability such that the entrepreneur always opts to undertake production once his fixed cost has been sunk. In this sense, occupational choice is irreversible in the model.

Suppose that all individuals (including landlords) have identical preferences represented by the indirect utility function $V(p, y)$, where y stands for the appropriate income variable. For algebraic convenience, let us further assume that the function takes the particular form $V(p, y) = h(p)y^{1-\gamma}/(1-\gamma)$ for $\gamma > 0$, $\gamma \neq 1$, or $V(p, y) = h(p)\log y$ for $\gamma = 1$. The indirect utility function has this form if underlying preferences are homothetic and individuals exhibit constant relative aversion to income risk (γ is the Arrow-Pratt measure of relative risk aversion, as extended to a many-commodity setting by Stiglitz, 1969).[5] In equilibrium, neither the entrepreneurial activity nor the labor activity can be strictly preferred, if identical agents face a choice between the two and a positive number select each occupation. Incomplete specialization thus requires that the expected utility of an entrepreneur be equal to the utility derived from the (certain) wage earned by a laborer, or that

$$(3) \qquad E\left\{h(p)\left[pF(l_z^i, \alpha^i) - wl_z^i\right]^{1-\gamma}/(1-\gamma)\right\} = h(p)w^{1-\gamma}/(1-\gamma),$$

[3]The analysis could also be carried out under the assumption that entrepreneurs hire labor prior to the resolution of uncertainty, rather than afterwards. The main results of this paper do not depend upon which of these alternative timing assumptions is made.

[4]In what follows, I shall also assume that N is sufficiently large relative to the size of the sector so that aggregate sectoral risk is negligible.

[5]Many, but not all, of the results of the paper go through without this assumption, which, however, greatly simplifies the exposition.

where l_z^i in (3) is now the optimal (state-dependent) choice of labor input, as determined by (2). Since p is nonstochastic, (3) can be written more simply as

$$(4)\quad E\left\{\left[pF\left(l_z^i,\alpha^i\right)-wl_z^i\right]^{1-\gamma}\right\}=w^{1-\gamma}.$$

The model is closed by the labor-market-clearing condition. At the time that production takes place the supply of labor is $L-N$, that is, the nonlandlord population less those who have chosen entrepreneurship as their profession. Full employment requires

$$(5)\qquad L_x+NEl_z^i=L-N.$$

Equations (1), (2), (4), and (5) determine the endogenous variables L_x, N, l_z^i, and w as functions of the exogenous variables, p and T, and the distribution and ultimate realization of α^i.

To recapitulate, I have assumed that production in the modern sector involves risk. Examples might include uncertainty about whether a new production process is viable or whether a particular manager is capable of organizing his employees efficiently. Such uncertainty is specific to the individual enterprise, but there are no institutions such as stock markets in the economy under consideration that would allow risks to be pooled by entrepreneurs, or shared by the various income classes. Entrepreneurship involves sufficiently large fixed costs, so that occupational choice is irreversible. Finally, hiring decisions are taken *ex post*, and profit maximization governs interfirm and intersectoral resource allocation.

It is easy to show that relative to a first-best allocation (which would result if markets were complete or if the government could plan production and distribute income), the equilibrium described above has too few entrepreneurs and exhibits overspecialization in agriculture. Note first that with independent entrepreneurial risks and firms that are sufficiently small relative to the size of the industrial sector, the economy as a whole is subject to no aggregate risk. Hence the first-best allocation is the one that maximizes the value of national output at international prices. The central planner's problem is to

$$(6)\qquad \max_{N,L_x,l_z^i}\ g(L_x,T)+pNEf\left(l_z^i,\alpha^i\right)$$

$$\text{subject to}\quad L_x+NEl_z^i=L-N.$$

Let tildes denote variables in the first-best situation. Manipulation of the first-order conditions to (6) gives

$$E\left\{pf\left(\tilde{l}_z^i,\alpha^i\right)\right\}-g_L(\tilde{L}_x,T)E\tilde{l}_z^i=g_L(\tilde{L}_x,T)$$

or, if we substitute $\tilde{w}\equiv g_L(\tilde{L}_x,T)$ (where $\tilde{w}$ is the shadow price of labor at the optimum),

$$(7)\qquad E\left\{pf\left(\tilde{l}_z^i,\alpha^i\right)-\tilde{w}\tilde{l}_z^i\right\}=\tilde{w}.$$

By comparison, application of Jensen's inequality to (4), recalling the fact that γ is positive, implies $E\{pf(l_z^i,\alpha^i)-wl_z^i\}>w$. It follows immediately that $\tilde{w}>w$.

Now, the marginal products of labor in agriculture in the equilibrium and in the first-best allocation are equal to the market and shadow wage rates, respectively. Thus, $\tilde{w}>w$ implies $\tilde{L}_x<L_x$ and $E\tilde{l}_z^i<El_z^i$. These in turn imply $\tilde{N}>N$. Overspecialization in agriculture and a shortage of entrepreneurs are direct consequences of the absence of risk-sharing markets. The question that arises, then, is whether or not this inefficiency in risk bearing and the attendant resource misallocation are exacerbated by the presence of free international trade or direct foreign investment.

II. Free Trade vs. Autarky

The size of the entrepreneurial class and the intersectoral allocation of resources will, of course, be different in the free-trade and autarky equilibria, because the relative price of the industrial sector good is altered by the presence of international trade. It is useful to begin by considering the effects on equilibrium allocations, incomes, and welfare levels of an exogenous change in the relative commodity price. Then the comparison of the alternative equilibria can be conducted by integrating the price derivatives between the autarky and free-trade price levels.

The effects of a relative price change are straightforward to derive. First, divide both sides of equation (4) by $p^{1-\gamma}$, and note that this equation determines a unique value for the real product wage in terms of the industrial good. The supply of entrepreneurs is perfectly elastic at this real product wage, and incomplete specialization, which requires a positive supply of both entrepreneurs and labor, can only occur if this particular value for w/p is realized. Letting a circumflex denote a proportional derivative, it follows from (4) that

$$(8) \qquad \hat{w} = \hat{p}.$$

Then, from (2), we have

$$(9) \qquad dl_z^i = 0.$$

Differentiation of (1), and substitution of (8) gives

$$(10) \qquad \hat{L}_x = -\sigma_x/(1-\theta_{Lx})\hat{p},$$

where σ_x is the elasticity of substitution between land and labor in agriculture, and θ_{Lx} is the labor share in agricultural income. The proportional change in the rental rate, r, is found by differentiating the price-equals-unit-cost equation for agriculture, and is given by

$$(11) \qquad \hat{r} = -\theta_{Lx}/(1-\theta_{Lx})\hat{p}.$$

Finally, from (5) and (8) through (10),

$$(12) \qquad \hat{N} = \frac{\sigma_{Lx}\lambda_{Lx}}{(1-\lambda_{Lx})(1-\theta_{Lx})}\hat{p},$$

where λ_{Lx} is the fraction of the nonlandlord population employed in the agricultural sector.

What do these results imply about the effects of free trade? First, equations (9), (10), and (12) together show that supply responses are "normal" in the model; that is, an increase in the relative price of the z good causes an increase in industrial output and a fall in agricultural output. Preferences are, by assumption, identical and homothetic, so aggregate demands must be negatively related to own-prices. Hence, the economy's excess demand for the industrial sector good is everywhere a nonincreasing function of the relative price of this good. It follows that if the economy is an importer of this good, the autarky price must lie above the free-trade price. This is likely to be the case, both because the less developed country might suffer from a technological disadvantage in modern sector production, and because the existence of stock markets which facilitate risk sharing in the more advanced countries implies that any bias against industrial output caused by the market imperfection will be quantitatively less important there.

Let us assume that the *LDC* does import the industrial sector good in the free-trade equilibrium. Then from (12) it is clear that "openness" indeed can be held responsible for inhibiting the formation of the local entrepreneurial class. Furthermore, in the absence of any government redistribution policy, free trade has a striking and unambiguous effect on the welfare levels of the various income classes in the *LDC* economy. Landlords necessarily benefit from free trade (relative to autarchy), because as p falls, the real rental rate, measured in terms of either good, rises. Similarly, the labor class is harmed by the introduction of trade, since the wage remains constant in units of the industrial sector good, but falls in terms of agricultural products. Finally, entrepreneurs must lose from trade as well, because equation (3) ties their level of expected utility to the welfare level of the laborers. These results are, of course, reminiscent of the Stolper-Samuleson findings for the familiar Heckscher-Ohlin model of trade. The similarity stems from the fact that the *ex ante* transformation schedule relating the number of entrepreneurs to the number of laborers is perfectly elastic at a particular product real wage. When this real wage prevails, the nonlandlord population is "as if" a single factor of production.

In circumstances where the introduction of free trade generates both "winners" and "losers," it is natural to ask whether or not a feasible compensation scheme exists that could guarantee a Pareto welfare improvement. This standard gains-from-trade ques-

tion is especially interesting in the present context, because we have noted that free trade effects a contraction of the entrepreneurial class, and an expansion of the agricultural sector, from an initial autarky situation in which the former is smaller, and the latter is larger, than would be first-best optimal.

The redistributive policy that I shall consider is a proportional income tax-cum-subsidy scheme, with tax at rate τ on rental income and subsidy at rate s on "earned" income. Implementation of these policy instruments is more likely to be feasible in *LDC*s than would be lump sum taxes and subsidies, and unlike the latter policies, proportional income taxes do not alter the allocation of resources or occupational-choice decision under the assumption of constant relative risk aversion. My strategy will be to construct a tax scheme under free trade that restores all income classes to their autarky level of welfare, and then check whether positive government revenue is thereby created. If so, it would be possible to lower a tax or raise a subsidy slightly to generate a situation that Pareto-dominates autarky.

The utility of landlords, U^T, is given by

$$U^T = h(p)[r(1-\tau)]^{1-\gamma}/(1-\gamma)$$

from whence (applying Roy's identity),

$$(13)\quad dU^T = h(p)[r(1-\tau)]^{-\gamma} \times\{-c^T dp + (1-\tau)\,dr - r\,d\tau\},$$

where c^T is the consumption of industrial sector goods by landlords. As the relative price moves from its autarky to its free trade level, the welfare of landlords can be held constant by continuous variation of the income tax rate that satisfies (noting (11))

$$(14)\quad d\tau|_{dU^T=0} = -(1-\tau)\left[\beta + \frac{\theta_{Lx}}{1-\theta_{Lx}}\right]\hat{p}$$

where β is the proportion of expenditure devoted to z goods.

Similarly, the utility of laborers, U^L, can be maintained at a constant level by an income subsidy that varies with the relative price change according to

$$(15)\quad ds|_{dU^L=0} = (\beta-1)(1+s)\hat{p}.$$

Recall that the expected utility of entrepreneurs is equal to that of laborers, so that (15) also fixes the welfare of entrepreneurs.

Finally, the government budget surplus is

$$B = \tau r - s(x + pz - r)$$

so that

$$dB = (s+\tau)r\hat{r} + r(ds + d\tau) - (x + pz)\,ds - spz\hat{p} - s(dx + p\,dz).$$

Substitution of (10), (14), and (15) into (16), and some simple manipulation, yields

$$(17)\quad dB = p(z-c)\hat{p} - s(dx + p\,dz),$$

where c is aggregate consumption of the z good. The first term is clearly positive, since both $\hat{p}$ and exports of the industrial sector good are negative. The second term can be computed from $dx = w\,dL_x$, and (with $dl_z^i = 0$),

$$dz = [Ef(l_z^i, \alpha^i)]\,dN = -\frac{Ef(l_z^i,\alpha^i)}{1+El_z^i}\,dL_x.$$

Thus,

$$-s(dx + p\,dz) = \frac{s(E\pi^i - w)}{1+El_z^i}\,dL_x,$$

where $E\pi^i$ is the expected income (i.e., profits) of the representative entrepreneur. This term is positive as well, since $E\pi^i > w$ by risk aversion, and both s and dL_x remain positive as p falls.[6]

Despite the second-best setting created by the absence of risk-sharing markets, the occupational choice model of entrepreneurship generates potential gains from trade. A sys-

[6] Similar reasoning establishes gains from trade also in cases where the economy exports the industrial sector good. Then $(z-c)$ and $\hat{p}$ in (17) would be positive, and s would be negative, implying again $dB > 0$.

tem of proportional income taxes and subsidies can always be constructed such that the social classes which benefit from trade can compensate those that are harmed by it. Of course, whether or not the required redistribution will actually take place in any particular country is, as always, a political question that cannot be answered here. Nonetheless, the analysis serves to demonstrate that one cannot infer solely from the fact that free trade inhibits the development of the entrepreneurial class that trade is harmful.

III. Can Protection be Welfare Improving?

Although free trade with appropriate redistribution is necessarily better than no trade, policies that alter the intersectoral allocation of resources might conceivably raise social welfare above the free-trade level, if they can work so as to offset the distortion associated with the inefficient allocation of risk. Such "tariffs and production subsidies as insurance" have been studied in other contexts by Newbery and Stiglitz, and by Eaton and myself. However, the mere existence of risk under an incomplete market structure is not sufficient for the nonoptimality of free trade (see my forthcoming article), so it is not obvious a priori whether trade or industrialization policy can be beneficial in the present context.

Consider then a tax on agricultural output at *ad valorem* rate t. As before, let us assume that the change in regime is accompanied by the implementation of proportional income taxes and subsidies that preserve the initial (in this case, free-trade) levels of utility. It is easy to check that this policy package is fully equivalent to a production subsidy in industry plus the appropriate redistributive measures, so it will not be necessary to investigate the latter interventions separately.

A change in the tax on agricultural output does not affect the product real wage (in units of industrial sector goods), as determined by equation (3). Since p does not change (because terms of trade are fixed for a small country), neither does the wage rate. This implies that the welfare levels of laborers and entrepreneurs are unaffected by the policy change, and no direct taxes or subsidies to their incomes are needed. There remains, therefore, only to check whether or not the revenue generated by the production tax is sufficient to compensate the landlords for their income losses.

After-tax income of landlords is $r(1-\tau)$ where $r=(1-t)x-wL_x$. A constant level of utility requires that τ be varied to satisfy

$$r\,d\tau=-(1-\tau)x\,dt. \tag{18}$$

The corresponding change in the government budget (where $B=tx+\tau r$) is given by

$$dB=t\,dx+(1-\tau)x\,dt+r\,d\tau. \tag{19}$$

Substitution of (18) into (19) yields

$$dB=t\,dx. \tag{20}$$

The right-hand side of (20) is negative for noninfinitesimal changes in t about $t=0$, since dx/dt is clearly negative. But government deficits, as required here, are infeasible in the model. The implication is that any noninfinitesimal tax or subsidy to agricultural production must generate utility losses for some individuals in the economy.

This finding can be understood as follows. Intersectoral policy does indeed alter the supply of entrepreneurs; but it does not bring about a more efficient allocation of risk across agents. Since risk misallocation, rather than the number of entrepreneurs, is the economically relevant manifestation of the distortion created by the incomplete market structure, a policy instrument which merely shifts resources between sectors and occupations is bound to be a failure. Instead, policy intervention should be tailored to work at the distorted margin—a point that was emphasized by Robert Baldwin (1969) in his analysis of infant-industry protection. There, as here, the mere existence of a distortion is not enough to justify intervention with the blunt tools of trade policy.[7]

[7] Kanbur (1981) analyzes the welfare effects of alternative policy interventions in a one-sector model of entrepreneurship and occupational choice. An example of a policy which facilitates risk spreading, and thereby improves the efficiency of resource allocation, is a *progressive* income tax-cum-subsidy scheme.

Note, furthermore, that tariff intervention is even more costly here than is a production tax-cum-subsidy. The former entails consumption losses, which are in addition to the production losses implicit in (20) above.[8] Indeed, with homothetic preferences, a policy of free trade constitutes a type of constrained optimum for an economy with an endogenous entrepreneurial class. If the institutional constraints on risk sharing are taken as immutable, then the efficiency gains associated with specialization according to (endogenous) comparative advantage are the most that can be attained.

IV. Foreign Investment and Local Entrepreneurs

Foreign investment, even more so than international trade, has been criticized on the grounds that it inhibits the development of an entrepreneurial class. It has been argued that foreign enterprises merely crowd out local efforts, and thus impart few if any benefits on the *LDC* economy. It is easy to see, without any formal algebra, that the logic of this argument is essentially confirmed in this model of the formation of the entrepreneurial class.

Consider the effects of an inflow of foreign enterprises into the industrial good sector of the *LDC* economy. Managers of these foreign-owned establishments face an infinitely elastic supply of labor at the real wage determined by equation (3). Whatever their demand for labor, there will be no effect on the domestic wage rate, provided that some domestic businesses continue to operate in the cum-foreign-investment equilibrium. The upshot is that an inflow of foreign enterprises has no impact on the welfare of any group in the domestic economy. With wages unaffected, landlord rents are constant at their pre-foreign-investment levels, and domestic entrepreneurs continue to have expected utility equal to the (unchanged) level enjoyed by laborers. All the surplus derived from the establishment of the new businesses accrues to the foreigners.

Furthermore, the inflow of foreign enterprises does indeed have the effect of crowding out local ventures. The supply of domestic entrepreneurs must shrink so as to release the individuals needed to serve as workers in the foreign firms. Although this shift in the occupational distribution has no direct consequences for welfare, it does imply a decline in national income for the less developed country, since the wages earned by the (new) laborers are less than the sum of the profits of the (former) entrepreneurs. The fall in aggregate income is, essentially, a premium paid by the domestic economy in exchange for the income insurance provided by the foreigners.

It does not follow from this analysis that foreign investment ought to be prohibited. For one thing, a tax on foreign profits would allow the host country to share in any surplus created by the inflow of foreign firms. But more to the point, we have seen that the size of the entrepreneurial class is not, in and of itself, a sensible policy target for the less developed economy. The contraction of the supply of local entrepreneurs when faced with competition from abroad—whether in the form of international trade or of direct foreign investment—should be viewed as symptomatic of a more fundamental market failure, namely the inability of the economy to share its production risks in an efficient manner. Policy efforts should be devoted to rectifying this inefficiency, rather than reacting to its consequences.

V. Conclusions

In this paper I have examined the argument that free trade and foreign investment may be harmful to less developed countries, because such international competition inhibits the formation of a local entrepreneurial class. A two-sector model of a small, open economy was developed in which the size of the entrepreneurial class is endogenously determined. Following Kanbur (1979, 1981), I viewed the entrepreneur as the manager of

[8] If preferences were not homothetic, then the consumption tax element of a tariff together with proportional redistribution of the tariff proceeds would allow redistribution of the entrepreneur's income across the firm-specific states of nature. In this case, the tariff would provide some insurance to the individual entrepreneur, and some trade intervention could be welfare improving.

the industrial enterprise, as well as the agent who bears the risks associated with industrial production in an economy in which markets for risk sharing are absent. In the model, the nonlandlord population faces an *ex ante* occupational choice decision between the entrepreneurial and labor activities.

It was shown that the size of the entrepreneurial class so determined is smaller than would be first-best optimal in the presence of efficient risk-sharing institutions such as stock markets. Free international trade causes the supply of local entrepreneurs to fall relative to autarky, if the *LDC* imports the industrial sector good in equilibrium. Nonetheless, there are potential gains from trade, in the sense that a feasible scheme of income taxes-cum-subsidies can always be devised to allow those who benefit from trade (landlords) to compensate those who lose (entrepreneurs and laborers).

Protectionist policies, such as tariffs on industrial sector imports or taxes on agricultural output, can be used to augment the size of the entrepreneurial class. However, these policies will always have deleterious welfare consequences. It turns out, much as in Baldwin's case against infant-industry protection, that intersectoral policy instruments are too blunt to deal with the distortion *implicit* in the absence of risk-sharing institutions. To improve allocative efficiency, an intervention would need to provide insurance to potential entrepreneurs against adverse outcomes in their industrial sector ventures. Trade policy does not satisfy this criterion.

An inflow of foreign-owned industrial enterprises into the less developed country also has an adverse effect on the supply of local entrepreneurs. In the absence of any tax on the profits of these firms, all the surplus from direct foreign investment accrues to the foreigners. The less developed country experiences a decline in national income, but the (expected) utility levels of all social classes in the economy remains fixed at pre-investment levels.

An important lesson that emerges from the analysis is that the size of the risk-bearing entrepreneurial class should not, in and of itself, be a policy target in less developed economies. Rather, policy should aim to provide the mechanisms by which risk can be efficiently allocated across the population.

REFERENCES

Baldwin, Robert E., "The Case Against Infant-Industry Tariff Protection," *Journal of Political Economy*, May 1969, *77*, 295–305.

Bhagwati, Jagdish N., "The Generalized Theory of Distortions and Welfare," in his et al., eds., *Trade, Balance of Payments and Growth*, Amsterdam: North-Holland, 1971.

Eaton, Jonathan and Grossman, Gene M., "Tariffs as Insurance: Optimal Commercial Policy when Domestic Markets are Incomplete," Working Paper No. 797, National Bureau of Economic Research, November 1981.

Grossman, Gene M., "The Optimal Tariff for a Small Country Under International Price Uncertainty: A Comment," *Oxford Economic Papers*, forthcoming.

Hirschman, Albert O., "How to Divest in Latin America, and Why," Essays in International Finance No. 76, International Finance Section, Princeton University, November 1969.

Kanbur, S. M., "Of Risk Taking and the Personal Distribution of Income," *Journal of Political Economy*, August 1979, *87*, 769–97.

______, "Risk Taking and Taxation: An Alternative Perspective," *Journal of Public Economics*, April 1981, *15*, 163–84.

Kihlstrom, Richard E. and Laffont, Jean-Jacques, "A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion," *Journal of Political Economy*, August 1979, *87*, 719–48.

Kilby, Peter, *Entrepreneurship and Economic Development*, New York: Free Press, 1971.

Lewis, W. Arthur, *The Theory of Economic Growth*, London: George Allen and Unwin, 1955.

Newbery, David M. G. and Stiglitz, Joseph E., *The Theory of Commodity Price Stabilization*, Oxford: Oxford University Press, 1981.

Pazos, Felipe, "The Role of International

Movements of Private Capital in Promoting Development," in J. H. Adler, ed., *Capital Movements and Economic Development*, New York: St. Martins Press, 1967.

Rostow, Walt W., "The Take-off Into Self-Sustained Growth," *Economic Journal*, March 1956, *66*, 25–48.

Stiglitz, Joseph E., "Behavior Toward Risk with Many Commodities," *Econometrica*, October 1969, *37*, 660–67.

A Test for Subadditivity of the Cost Function with an Application to the Bell System

By DAVID S. EVANS AND JAMES J. HECKMAN*

The recent literature on multiproduct firms demonstrates that when all firms in an industry (actual or potential) have access to a common technology, properties of the firm cost functions reveal the most efficient industry structure (see, for example, Elizabeth Bailey and Ann Friedlaender, 1982). An important issue addressed in this literature is the derivation of conditions which guarantee that an industry is a natural monopoly. Despite the relevance of this issue to discussions concerning the desirability of competition in regulated industries, few empirical studies offer reliable evidence on this question. The reason for this is that the theory of multiproduct industries (see, for example, William Baumol, John Panzar, and Robert Willig, 1982) suggests that global information about cost functions is required to determine the presence of natural monopoly. Such information is seldom available.

This article proposes a new test of necessary conditions for natural monopoly that does not require global information on firm cost functions. Our test does not require the extrapolation of estimated cost functions well outside the range of the available data that are required in tests currently proposed in the literature.

To see why global information is required to test for necessary *and* sufficient conditions for natural monopoly, consider a firm which produces an output vector $q=(q_1,\ldots,q_n)$ with cost function $C(q)$. A firm is a natural monopoly if and only if $C(q)$ is subadditive over the relevant range of output levels.[1,2] The function $C(q)$ is subadditive at the output level $\tilde{q}$ if and only if

$$(1) \qquad C(\tilde{q}) < \sum_{i=1}^{n} C(\tilde{q}^i),$$

where for nonnegative $\tilde{q}^i$

$$(2) \qquad \sum_{i=1}^{n} \tilde{q}^i = \tilde{q},$$

with at least two vectors $\tilde{q}^i$ nonzero, for all $\tilde{q}^i$ satisfying (2). Thus an industry is a natural monopoly if a single firm can produce all relevant output vectors more cheaply than two or more firms. To test this condition requires knowing the cost of all output vectors $\tilde{q}^i$ smaller than $\tilde{q}$ and thus requires global information about the cost function. Such knowledge about $C(q)$ is required for all possible industry equilibrium output values.

Less empirically demanding necessary and sufficient conditions for subadditivity remain to be developed. Baumol et al. derive separate necessary and sufficient conditions which require less information than the joint necessary and sufficient conditions for subadditivity.[3] For example, the presence of economies of scope is a necessary condition

*Departments of Economics, Fordham University, Bronx, NY 10458 and University of Chicago, Chicago, IL 60637, respectively. We are indebted to George Yates for programming the subadditivity test, Thomas Coleman and John Bender for research assistance, and Kevin M. Murphy, Rodney Smith, and Lester Telser for helpful comments. The paper has benefited from seminar presentations at the Applied Price Theory Workshop at the University of Chicago and the Graduate School of Business, New York University.

[1]Subadditivity of the firm cost function is a necessary and sufficient condition for natural monopoly only when all firms have access to the same technology and when market coordination between separate firms is unable to achieve the same economies (say by networking or pooling arrangements) as internal coordination within a single firm. These assumptions may not characterize many real-world industries. See our (1983a) article for further discussion.

[2]The relevant range of outputs depends on the demand and cost conditions prevailing over the period of interest to the analyst or policymaker.

[3]See Baumol et al. or Bailey and Friedlaender for formal statements of these conditions.

for subadditivity. Economies of scope exist if the costs of producing each product separately exceed the costs of producing all products jointly. Economies of scope and declining average incremental cost for each product are sufficient conditions for subadditivity. The average incremental cost of output i, when the total output vector is $\tilde{q}$, is the cost of producing output vector $\tilde{q}$ less the cost of producing all of $\tilde{q}$ excluding product i divided by the output of $i, \tilde{q}_i$. The average incremental cost for product i declines if the average incremental cost of producing i decreases with increases in the level of $\tilde{q}_i$.[4]

Baumol et al. recommend testing the necessary and sufficient conditions for subadditivity separately. If the necessary condition is rejected, subadditivity is rejected. If the sufficient conditions are accepted, subadditivity is accepted. Unfortunately, even this restrictive procedure for testing subadditivity requires more information than is usually available for most industries. Their test requires observations on the stand-alone costs of production for each product in the output vector and observations on the costs of production for all output vectors containing positive quantities of $n-1$ outputs and a zero quantity of the other output.

To illustrate the problems that arise in implementing their test, consider testing for subadditivity in the postal industry. This industry provides several types of mail service, some of which have been opened to competition in recent years. It would be interesting to determine whether a single firm could provide all postal service more cheaply than could two or more firms. But data on the stand-alone costs of providing first-class mail service are not available. Consequently, calculating economies of scope between first-class and other postal services requires extrapolation of estimated cost functions far outside the range of observations over which they are estimated.[5] Similar problems beset the calculation of average incremental costs. Moreover, the Baumol et al. test may prove inconclusive even when data are available for calculating reliable estimates of the relevant quantities. Acceptance of their necessary condition but rejection of their sufficient conditions may occur.

This article proposes a new test for subadditivity within a region that avoids the need to extrapolate outside the range of the available data. The test is local and not global. It is based on the idea that if subadditivity is rejected in one region, global subadditivity must be rejected. Section I describes our test for a two-product industry. (Generalization to n-product industries is straightforward.) Section II applies the test to data from the U.S. Bell System.

I. Subadditivity Test

We begin by refining the terminology used in the introduction. Consider a two-product industry in which all firms have access to the same technology. The cost function $C(q_1, q_2)$ is subadditive at $\tilde{q} = (\tilde{q}_1, \tilde{q}_2)$ if and only if for nonnegative $\tilde{q}_1, \tilde{q}_2$

$$(3) \qquad \Sigma_i C(a_i\tilde{q}_1, b_i\tilde{q}_2) > C(\tilde{q}_1, \tilde{q}_2), \qquad i = 1, \ldots, n,$$

$$(4) \quad \Sigma a_i = 1, \quad \Sigma b_i = 1, \quad a_i \geq 0, \quad b_i \geq 0,$$

for at least two a_i or b_i not equal to zero. It is superadditive at $\tilde{q}$ if and only if " $>$ " is replaced with " $<$ " in (3). It is additive at $\tilde{q}$ if and only if " $>$ " is replaced with " $=$ " in (3). A firm with superadditive costs could save money by breaking itself into two or more divisions. Unless there are firm-specific fixed factors which preclude such decentralization, we would not expect to observe profit-maximizing firms operating with superadditive costs. A firm with additive costs

[4] Economies of scope and average incremental costs can also be defined for subsets of the output vector containing multiple products. See Baumol et al.

[5] Melvyn Fuss and Leonard Waverman (1981) have used this test to determine whether Bell Canada has a natural monopoly over local, toll, and private line telephone services. They reject the hypothesis that there are economies of scope between private line service and local and toll service. But, as they readily admit, their test is unreliable since they have to extrapolate the cost function far outside their sample in order to calculate stand-alone costs for local, toll, and private line services.

may have decentralized itself into the optimal configuration. Consequently, in many situations the interesting statistical question concerns whether the cost function is additive or subadditive at observed output levels. The cost function is quasi-globally subadditive, superadditive, or additive if and only if the cost function is subadditive, superadditive, or additive, respectively, at all relevant output vectors $\tilde{q}$. The relevant output vectors are those which are consistent with industry equilibrium given demand and cost conditions for alternative possible organization patterns of the industry (for example, multifirm vs. single firm).

Our test computes (3) for an "admissible range" of outputs. For simplicity, we restrict our evaluation of (3) to $n=2$ so that we compare two-firm configurations with the monopoly configuration. We also confine our attention to the case where the analyst has time-series data on a single firm assumed to be a monopoly, although the test can be readily applied to situations where the analyst has cross section or panel data on a sample of firms. Denote each time-series observation by an output q_t, $t=1,\dots,T$.

Denote the first hypothetical firm by A and the second hypothetical firm by B. Consider output level $\tilde{q}$ and let $\tilde{q}^A+\tilde{q}^B=\tilde{q}$. Figure 1 illustrates our test. We define an "admissible region" represented by an area on the floor of the diagram. The cost of producing $\tilde{q}^A$ is $\tilde{C}^A$, the cost of producing $\tilde{q}^B$ is $\tilde{C}^B$, and the cost of producing $\tilde{q}$ is $\tilde{C}$. If $\tilde{C}^A+\tilde{C}^B>\tilde{C}$, then the cost function is subadditive at $\tilde{q}$ with respect to the particular two-firm industry configuration described by $(\tilde{q}^A,\tilde{q}^B)$, a situation which holds true in Figure 1. We consider all two-firm configurations $(\tilde{q}^A,\tilde{q}^B)$ in the admissible region which sum to $\tilde{q}$. If $\tilde{C}^A+\tilde{C}^B>\tilde{C}$ for all such configurations, then the cost function is subadditive at $\tilde{q}$ over the admissible region, the situation depicted in Figure 1.

Our choice of an admissible region for each output $\tilde{q}$ is dictated by our desire to avoid "excessive" extrapolation outside the data. It is difficult to make this notion precise without knowing in advance exactly what it is we seek to estimate. Approximation theory provides bounds for particular ap-

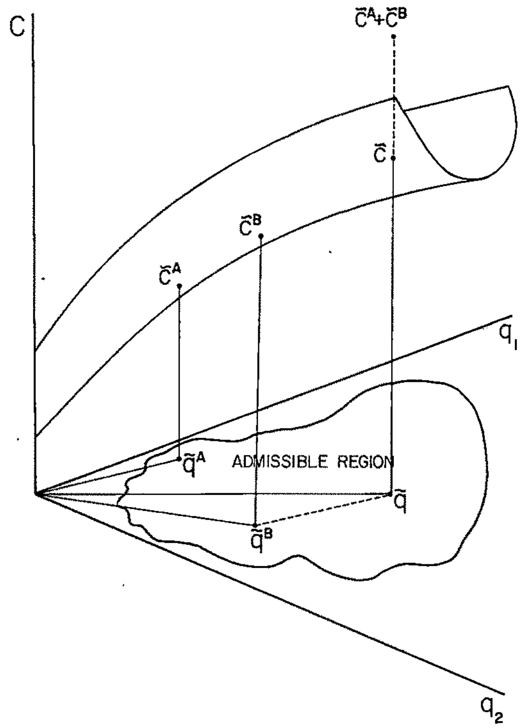


FIGURE 1. TEST FOR SUBADDITIVITY

proximations to known functions. But these bounds are of little use in applied work since, in practice, the true cost function is unknown.[6] In this paper we define the admissible region to keep hypothetical industry output configurations within the range of output configurations actually observed in the data.

Specifically, we define the admissible region so that it satisfies two constraints. The first is that no hypothetical firm in the two-firm industry is permitted to produce less of either output than the firm for which we have data. Define q_M as the vector of minimal sizes of the output of the firms permitted by this criterion:

$$q_M=(q_{1M},q_{2M})=\left(\min_t q_{1t},\ \min_t q_{2t}\right).$$

[6]See, for example, Philip Davis (1975) for a discussion of approximation theory and bounds for various functions.

Firm A produces

$$(5) \quad q_t^A = \left(\phi q_{1t}^* + q_{1M}, \omega q_{2t}^* + q_{2M}\right).$$

Firm B produces

$$(6) \quad q_t^B = \left((1-\phi) q_{1t}^* + q_{1M}, (1-\omega) q_{2t}^* + q_{2M}\right).$$

The parameters (ϕ, ω) satisfy

$$(7) \quad 0 \le \phi \le 1, 0 \le \omega \le 1.$$

Aggregating across firms A and B, we obtain

$$(8) \quad q_{1t}^* + 2q_{1M} = \tilde{q}_{1t}, \quad q_{2t}^* + 2q_{2M} = \tilde{q}_{2t},$$

so that

$$(9) \quad q_{1t}^* = \tilde{q}_{1t} - 2q_{1M}, \quad q_{2t}^* = \tilde{q}_{2t} - 2q_{2M},$$

for $\tilde{q}_{it} > 2q_{iM}$. We restrict the test to values of $\tilde{q}_t$ which satisfy this inequality. We test in years for which the output of both goods is at least twice the lowest output level in the sample.

The second constraint is that both firm A and B produce q_1 and q_2 in a ratio within the range of ratios observed in the data. Thus we require

$$(10) \quad R_L < \frac{\phi q_{1t}^* + q_{1M}}{\omega q_{2t}^* + q_{2M}} < R_U,$$

$$R_L < \frac{(1-\phi) q_{1t}^* + q_{1M}}{(1-\omega) q_{2t}^* + q_{2M}} < R_U,$$

where

$$(11) \quad R_L = \min_t q_{1t}/q_{2t}, \quad R_U = \max_t q_{1t}/q_{2t}.$$

This constraint precludes our hypothetical firms from specializing in either output to a greater extent than has the firm for which we have data.[7]

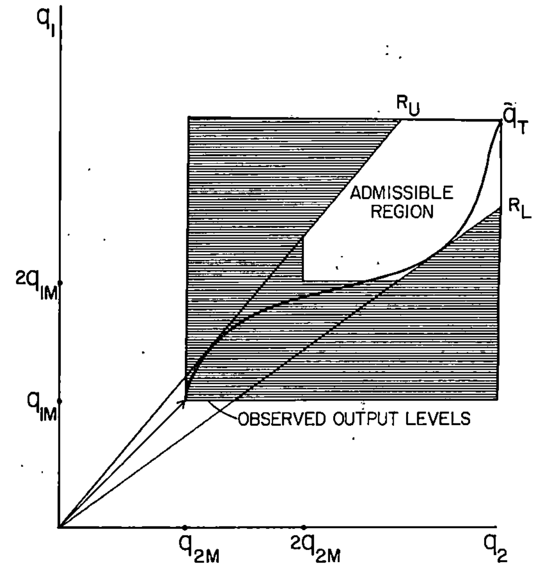


FIGURE 2. DETERMINATION OF ADMISSIBLE REGION

Figure 2 illustrates how these constraints restrict the admissible sample region. The smallest hypothetical firm size we consider is q_M. The hypothetical firm must also have output ratios between the vectors R_L and R_U, where R_L is the lowest ratio of output 1 to output 2 in the sample and R_U is the highest ratio of output 1 to output 2 in the sample. In addition, the test only applies to $\tilde{q}_t$ which satisfy (9) and therefore lie in the northeast quadrant of the box drawn in Figure 2.

[7]It is possible to define a broader admissible region by also including output configurations that are not too far removed from observed configurations. For example, we could modify the first constraint so that neither hypothetical firm is allowed to produce less than $1/\alpha$ of the minimum amount of output observed in the sample or more than α of the amount of output observed in the sample, where $\alpha \ge 1$. We could modify the second constraint so that the range of specialization for hypothetical firm lies between R_L/β and βR_U where $\beta \ge 1$. The admissible region described in the text is for the special case where $\alpha = \beta = 1$. We have chosen a conservative admissible region for two reasons. First, since this study reports the first application of our test for subadditivity, we believe it is best to be conservative and restrict the test to observed output combinations. Second, the estimates reported below deteriorate as we get further from the sample mean. Expanding the admissible region, at least for our example, does not provide much additional information. We would like to thank the referee for the suggestion for expanding the admissible region.

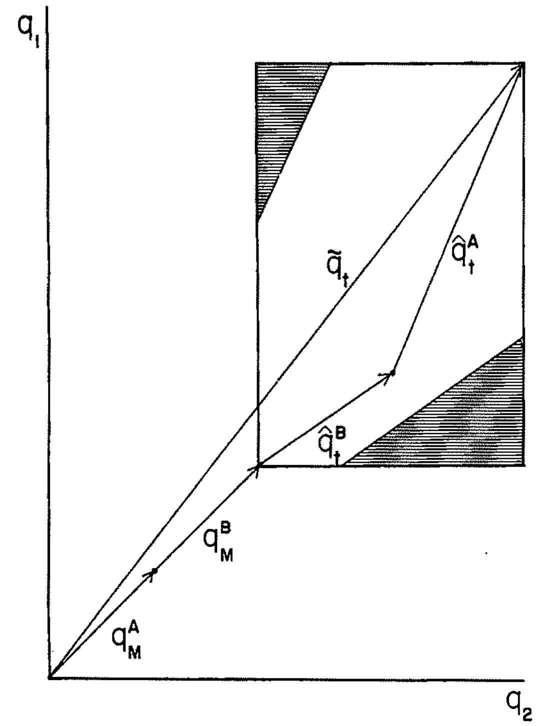


FIGURE 3. IMPLEMENTATION OF SUBADDITIVITY TEST

Figure 3 illustrates an implementation of the test for output level $\tilde{q}_t$ in the relevant output set. Let $\hat{q}_t^A=(\phi q_{1t}^*, \omega q_{2t}^*)$. Firm A produces q_M plus an increment $\hat{q}_t^A$ and firm B produces q_M plus an increment $\hat{q}_t^B$ where $\hat{q}_t^A$ and $\hat{q}_t^B$ are chosen so that $\hat{q}_t^A+\hat{q}_t^B+2q_M$ sum to the output level $\tilde{q}_t$ at which the test is being performed. All possible vectors $\hat{q}_t^A$ and $\hat{q}_t^B$ within the admissible region, and thus all possible two-firm output configurations within the admissible region, are evaluated. In practice, the $\hat{q}_t^A$ and $\hat{q}_t^B$ are constructed by varying ϕ and ω over a grid given by $\phi=i/g$, $\omega=k/g$, $i,k=0,\dots,g$ where g is the grid size.

We now formally describe our test for subadditivity. Let

$$\tilde{C}_t^A(\phi,\omega)=\tilde{C}(\tilde{q}_t^A)=\tilde{C}(q_M+\hat{q}_t^A),$$

$$\tilde{C}_t^B(\phi,\omega)=\tilde{C}(\tilde{q}_t^B)=\tilde{C}(q_M+\hat{q}_t^B),$$

$$\tilde{C}_t=\tilde{C}(\tilde{q}_t^A+\tilde{q}_t^B)=\tilde{C}(\tilde{q}_t).$$

We measure the degree of subadditivity by

$$Sub_t(\phi,\omega)$$
$$=\left[\tilde{C}_t-\tilde{C}_t^A(\phi,\omega)-\tilde{C}_t^B(\phi,\omega)\right]/\tilde{C}_t.$$

If $Sub_t(\phi,\omega)$ is less than zero, the cost function is subadditive with respect to the industry configuration given by (ϕ,ω). If $Sub_t(\phi,\omega)$ is equal to zero, the cost function is additive with respect to the industry configuration given by (ϕ,ω). If $Sub_t(\phi,\omega)$ is greater than zero, the cost function is superadditive over the industry configuration given by (ϕ,ω).

We calculate $\text{Max}_{(\phi,\omega)}Sub_t(\phi,\omega)$. If this quantity is negative and statistically significantly different from zero, we reject the hypothesis that the cost function is additive at $\tilde{q}_t$ over the admissible region. Values of the test statistic that are negative and statistically significantly different from zero do not contradict the hypothesis that the cost function is subadditive at $\tilde{q}_t$ relative to the admissible region. If we do not reject the composite hypothesis that the cost function is subadditive at all feasible output levels, then we do not reject the hypothesis that the cost function is quasi-globally subadditive relative to these output levels. A test of the hypothesis of subadditivity over the whole sample is based on $\text{Max}_{(\phi,\omega,T^*)}Sub_t(\phi,\omega)$, where T^* is the set of sample years for which feasible partitions exist.[8] Note that our criterion tests only necessary conditions for subadditivity as long as the admissible region is a proper subset of the possible

[8]Note that because the maximization is with respect to a continuous set of exogenous explanatory variables and because in general there is only one stationary point for the maximum problem, our procedure does not create an order-statistics problem, and so we can use conventional test statistics. We make no claim about the optimality of our test. A better test, that grew out of discussions with Kevin M. Murphy, computes $Sub_t(\phi,\omega)$ for all points in the admissible region. This is an infinite dimensional statistic but its distribution can be derived. A one-sided test of the hypothesis that this function is everywhere negative can be constructed that utilizes more information than the test proposed in this paper. While this test is no worse than the one proposed in the text, it is also more complicated to compute, and is not developed here.

region of output configurations. Failure to find subadditivity within the admissible region is informative in rejecting the hypothesis of subadditivity; evidence supporting subadditivity within the admissible region obviously does not indicate support for that hypothesis outside the admissible region.

II. Subadditivity Tests for the U.S. Bell System

We use data on the Bell System from 1947–77 to test whether the Bell System had a subadditive cost function at the output levels produced during those years. We assume the Bell System produces local and toll telephone services with capital, labor, and materials. We estimate several alternative cost function specifications under alternative assumptions concerning the structure of the disturbances. Using likelihood ratio statistics, we find that the preferred specification is a general translog cost function with first-order autoregressive disturbances in the cost and factor share equations. The estimated cost function is monotonic and concave with respect to input prices in all years. The estimated own-price elasticities of demand for capital, labor, and materials are negative in all years. Therefore our estimates satisfy the conditions required of an economically valid cost function.[9] The Appendix reports the estimated cost function.

Between 1947 and 1977, cost increased more than fourteenfold, toll service increased almost fourteenfold, and local service increased fivefold. The smallest quantities of local and toll service were both produced in 1947. Output doubled by 1958 making this year the first feasible one for our test. Over the sample period, the ratio of local to toll varies between 0.5 and 1.3.

TABLE 1—MAXIMUM PERCENT GAIN FROM MULTIFIRM PRODUCTION VS. SINGLE-FIRM PRODUCTION[a]

Year	Percent Gain	Standard Error	ϕ	ω
1958	13	15	0.0	0.0
1959	20	14	0.0	0.2
1960	25	14	0.0	0.4
1961	25	14	0.0	0.6
1962	33	14	0.0	0.5
1963	40	15	0.0	0.5
1964	44	15	0.0	0.6
1965	48	16	0.0	0.6
1966	53	23	0.5	0.9
1967	58	23	0.6	0.7
1968	51	26	0.5	0.8
1969	50	30	0.3	0.9
1970	39	22	0.5	0.7
1971	36	21	0.4	0.6
1972	39	21	0.4	0.6
1973	41	20	0.4	0.6
1974	42	21	0.4	0.6
1975	45	20	0.4	0.6
1976	59	20	0.5	0.5
1977	51	19	0.5	0.5

[a]Entries equal Max $Sub_t \times 100$. A positive number indicates that multifirm production is more efficient than single firm production.

For each year $t = 1958,\ldots,1977$, we calculate $Sub_t(\phi, \omega)$ over a grid with a $g = 10$ for (ϕ, ω) lying in the admissible region. Table 1 reports the estimates of Max $Sub_t(\phi, \omega)$. Table 2 reports the estimate of $Sub_t(\phi, \omega)$ for $t = 1961$, a year near the center of our data. We find that Max Sub_t is always greater than and often statistically significantly different from zero for output configurations produced between 1958 and 1977. Therefore we reject the hypothesis that the Bell System's cost function is subadditive at any of these output levels. We also reject the hypothesis that the Bell System's cost function is quasi-globally subadditive over these output levels. The fact that Max $Sub_t(\phi, \omega)$ is never significantly less than zero and the fact that Sub_t is frequently positive suggest that our finding that the Bell System cost function is not quasi-globally subadditive is robust.

The frequent positive and sometimes statistically significant point estimates of Sub_t

[9]Like other economists who estimate translog cost functions, we require our cost function to be linear homogeneous in input prices and to have a symmetric Hessian matrix with respect to input prices. See, for example, Laurits Christensen and William Greene (1976); Fuss and Waverman; Friedlaender, Clifford Winston, and Kung Wang (1983). Unlike these analysts, we report tests of these restrictions. We resoundingly reject them. This phenomenon occurs in all other specifications of the model that we estimate. We conjecture that these restrictions would be rejected in other translog cost function studies. In order to make our estimates consistent with the translog estimates reported by other economists, we impose homogeneity and symmetry restrictions on the cost function.

TABLE 2—PERCENT GAIN OR LOSS FROM MULTIFIRM VS. SINGLE-FIRM PRODUCTION ALTERNATIVE INDUSTRY CONFIGURATIONS, 1961 DATA[a]

	$\phi =$										
$\omega =$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.0	8										
	(21)										
0.1	8	8									
	(19)	(20)									
0.2	9	8	8								
	(18)	(18)	(19)								
0.3	12	10	9	9							
	(16)	(17)	(18)	(19)							
0.4	15	13	10	9	9						
	(15)	(13)	(17)	(18)	(18)						
0.5	20	16	13	11	9	9					
	(14)	(15)	(16)	(17)	(18)	(18)					
0.6	25	21	17	14	11	10	9				
	(14)	(15)	(16)	(17)	(17)	(18)	(18)				
0.7			23	18	15	12	10	9			
			(16)	(16)	(17)	(18)	(18)	(18)			
0.8					20	16	12	10	8		
					(17)	(18)	(18)	(19)	(19)		
0.9							17	13	10	8	
							(19)	(19)	(20)	(20)	
1.0										10	8
										(21)	(21)

[a]Entries equal $Sub_{1961} \times 100$. A positive number indicates that multifirm production is more efficient than single-firm production.

Standard errors are reported in parentheses.

The numbers reported here are based on an autoregressive translog cost function reported in our (1983b) article.

suggest that the Bell System was not optimally decentralized during our sample period. This finding is consistent with anecdotal evidence on the Bell System. *Fortune Magazine* reported that William L. Weiss, head of the new operating company for the Midwest region, "thinks AT&T's marketing was too centralized with the result that managers became 'less creative and more dependent on the signal caller'" (June 27, 1983, p. 64). The Midwest operating company will reportedly operate as a decentralized confederation of the five companies from which it is to be constituted.

III. Conclusions

This article presents a new test for subadditivity of the cost function that requires less information on the cost function than does the test proposed by Baumol, Panzar, and Willig. The test provides an easily computed rejection criterion for the hypothesis of subadditivity. Applying this test to the Bell System using 1947–77 time-series data, we reject the hypothesis that the Bell System's cost function is subadditive at the output levels produced between 1958–77. We find limited evidence that the Bell System did not optimally decentralize itself during these years and was therefore operating inefficiently.

APPENDIX

We estimate the Bell System cost function using aggregate time-series data on the Bell System for 1947–77. We estimate a cost function rather than a production function in order to make our approach consistent with previous studies of the production characteristics of the telecommunications industry, and because the theoretical literature on natural monopoly relies on the cost function rather than the production function. We disaggregate outputs into local and long distance services and inputs into capital, labor,

and materials. As a proxy for technological change, we follow previous studies of the production characteristics of the Bell System (see, for example, Laurits Christensen, Diane Cummings and Philip Schoech, 1981; M. Ishaq Nadiri and Mark Shankerman, 1981; and Hrishikesh Vinod, 1976), by using an index of lagged research and development expenditures by Bell Laboratories.

Following conventional practice (see, for example, Christensen and William Greene, 1976), we use the translog approximation to the cost function and estimate a system consisting of the cost equation, the capital share equation, and the labor share equation using maximum likelihood techniques. We find that a specification with first-order autoregressive disturbances maximizes the likelihood function and purges the residuals of serial correlation.[10,11] The estimated cost function is reported in Table 3. This cost function is monotonically increasing and concave with respect to all input prices in all years. The own-price elasticities of demand for capital, labor and materials are negative in all years. The estimates thus satisfy the sufficient conditions for an economically valid cost function.[12] We reject the hypothesis that the cost function is separable in local and long distance output. Thus it is not valid to aggregate local and long distance telephone service into a single measure of telephone service and estimate a single-product cost function as is done in the Christensen et al., Nadiri-Shankerman, and Vinod studies. Our (1983b) article provides further details on our estimation procedures and results and on our data sources.

[10]We also estimate two alternatives to the translog cost function. The first alternative applies a Box-Cox transformation to the output variables. The second alternative applies a Box-Cox transformation to all right-hand side variables. We are unable to reject the hypothesis that the correct specification is translog. See our (1983b) article.

[11]We assume that output levels and factor prices are exogenous. Although both assumptions are questionable on a priori grounds for the telephone industry, using a Durbin (1954)-Wu (1973) test, we reject the hypothesis that output levels and input prices are endogenous.

[12]But see fn. 9.

TABLE 3—ESTIMATED COST FUNCTION

Parameter	Coefficient Estimate	Standard Error
Constant	9.054	(.005)
Capital	.535	(.008)
Labor	.355	(.007)
Toll	.260	(.309)
Local	.462	(.226)
Technology	−.193	(.086)
Capital^2	.219	(.024)
Labor^2	.174	(.027)
Capital-Labor	−.180	(.019)
Toll^2	−8.018	(2.170)
Local^2	−4.241	(1.314)
Local-Toll	11.663	(3.144)
Technology^2	−.176	(1.033)
Capital-Toll	.337	(.138)
Capital-Local	−.359	(.122)
Labor-Toll	−.179	(.083)
Labor-Local	.164	(.071)
Capital-Technology	.083	(.053)
Labor-Technology	−.057	(.047)
Toll-Technology	−1.404	(1.497)
Local-Technology	1.207	(1.431)
Autocorrelation Parameter for:		
Cost Equation	.187	(.105)
Share Equations	.712	(.094)

Notes:

Summary Statistics	R^2	Degrees of Freedom	Durbin-*h*
Cost Function	.9999	15	.65
Capital Share Equation	.9756	27	1.50
Labor Share Equation	.9835	27	1.37

Generalized Variance for System = 10.568

REFERENCES

Bailey, Elizabeth E. and Friedlaender, Ann E., "Market Structure and Multiproduct Industries," *Journal of Economic Literature*, September 1982, *20*, 1024–41.

Baumol, William, Panzar, John C. and Willig, Robert D., *Contestable Markets and the Theory of Industry Structure*, San Diego: Harcourt, Brace, Jovanovich, 1982.

Christensen, Laurits, Cummings, Diane and Schoech, Philip, "Econometric Estimation of Scale Economies in Telecommunications," Working Paper No. 8124, SSRI, University of Wisconsin-Madison, September 1981.

______ **and Greene, William,** "Economies of Scale in U.S. Electric Power Distribution," *Journal of Political Economy*, October 1976,

84, 655–76.

Davis, Philip J., *Interpolation and Approximation*, New York: Dover, 1975.

Durbin, J., "Errors in Variables," *Review of the International Statistical Institute*, 1954, *20*, 22–32.

Evans, David S. and Heckman, James J., (1983a) "Natural Monopoly," in D. S. Evans, ed., *Breaking Up Bell: Essays on Industrial Organization and Regulation*, New York: North-Holland, 1983, 127–56.

______ and ______, (1983b) "Multiproduct Cost Function Estimates and Natural Monopoly Tests for the Bell System," in D. S. Evans, ed., *Breaking Up Bell*, New York: North-Holland, 1983, 253–82.

Friedlaender, Ann, Winston, Clifford and Wang, Kung, "Costs, Technology, and Productivity in the U.S. Automobile Industry," *Bell Journal of Economics*, Spring 1983, *14*, 1–20.

Fuss, Melvyn and Waverman, Leonard, *The Regulation of Telecommunications in Canada*, Ontario: Economic Council of Canada, 1981.

Nadiri, M. Ishaq and Shankerman, Mark, "The Structure of Production, Technological Change and the Rate of Growth of Total Factor Productivity in the U.S. Bell System," in T. Cowing and R. Stevenson, eds., *Productivity Measurement in Regulated Industries*, New York: Academic Press, 1981.

Vinod, Hrishikesh, "Application of New Ridge Regression Methods to a Study of Bell System Scale Economies," *Journal of the American Statistical Association*, December 1976, *71*, 835–41.

Wu, De Min., "Alternative Tests For Independence Between Stochastic Regressors and Disturbances," *Econometrica*, December 1973, *41*, 733–50.

Fortune Magazine, "Ma Bell's Kids Fight for Position." June 27, 1983, *107*, 62–69.

Do Oligopolists Earn "Noncompetitive" Rates of Return?

By WILLIAM W. ALBERTS*

Today, many industrial economists seem to accept two propositions about the rate of return earned across markets in manufacturing and trade by the leading firms in these noncompetitively structured markets.[1]

PROPOSITION 1: *These rates vary directly with the degree to which the markets are concentrated.*[2]

Figure 1a shows a stylized version of this relationship for a particular study interval. In this figure *CR* designates the four-firm concentration ratio, *R* designates the rate of return to invested equity capital, and R_C designates the rate of return that would have prevailed for the interval in competitively structured markets viewed as about equal in riskiness to the markets under study.

PROPOSITION 2: *The rates of return between* R_1 *and* R_2 *in Figure* 1a *are not only higher than* R_C, *but high.*[3]

High, in effect, is defined in this proposition in either of two ways. Most commonly, it has been taken to mean: high enough to warrant remedial intervention of some sort by the state.[4] Recently, however, a growing minority of economists has urged that it be taken to mean instead: high enough to warrant intervention, provided the state can show that rates of return in excess of R_1 reflect "collusive" behavior by the leading firms and not cost advantages which these firms have over their leading rivals.[5]

Both meanings in turn reflect a third: high enough to imply a typical market price closer to p_M in Figure 1b than to p_C, where p_M is the price that would prevail if the leading firms maximized collective, current-period profits and p_C is the price that would prevail if collective, current-period profits approximated zero.[6]

Proposition 1 rests on a large body of empirical work. Proposition 2, however, does not; nor does it rest on any theoretical analysis. Industrial economists simply have intuited that there is a correspondence between the R_1R_2 segment in Figure 1a and the p_Mp^* segment in Figure 1b. Are there substantive grounds for the intuition? I argue that there are not. The rates of return that lie along the R_1R_2 segment are "competitive,"

*Graduate School of Business Administration, University of Washington, Seattle, WA 98195. I thank Donald Dewey, Alan Hess, Gailen Hite, Joseph Monsen, Edward Rice, George Stigler, and Dean Worcester for comments on various earlier versions of this paper.

[1]There is evidence that in the industrial sector most producers perceive themselves to be oligopolists in the markets in which they operate. See, for example, Dean Worcester (1967, ch. IV) and Robert Buzzell (1981).

[2]Most of the studies articulating this proposition are summarized in Leonard Weiss (1971; 1974), F. M. Scherer (1980), and William Shepherd (1979).

[3]There are statements, or recapitulations, of this proposition (all of which use the term high) in Joe Bain (1956, p. 203), Harold Demsetz (1974, p. 176), Donald Hay and Derek Morris (1979, p. 226), James Koch (1980, p. 194), Dennis Mueller (1977, p. 369), Sam Peltzman (1977, p. 252), Shepherd (1979, p. 270), Roger Sherman (1974, p. 333), and Weiss (1974, p. 231).

[4]For example, Title 1 of the Hart Bill (which was introduced in 1972 and is reproduced in Harvey Goldschmid et al. 1974, pp. 444–45), identifies R_1 (or, perhaps, an average of R_1 and R_2) as 15 percent and specifies that under the bill any firm which generated a rate of return this high over the interval 1968–72 could be presumed to possess unlawful monopoly power.

[5]Demsetz (1973, 1974) and Peltzman both have urged this second reading. Further, they have argued that in most concentrated markets, the facts *will* show that high rates of return reflect cost advantages.

[6]It is difficult to find *explicit* statements of this third reading of Proposition 2 (although it is not difficult to find assertions that high rates of return imply high prices—see, for example, Weiss, 1974, p. 232). When authors impute, or appear to impute, a goal of *periodic* profit maximization to a leading firm in an oligopolistic market, though, they implicitly embrace this reading. See, for example, Richard Leftwich (1976, ch. 12), Edwin Mansfield (1982, pp. 349–50), Paul Samuelson (1981, p. 482), Scherer (pp. 186–88, but *not* ch. 8), and Shepherd (ch. 14).

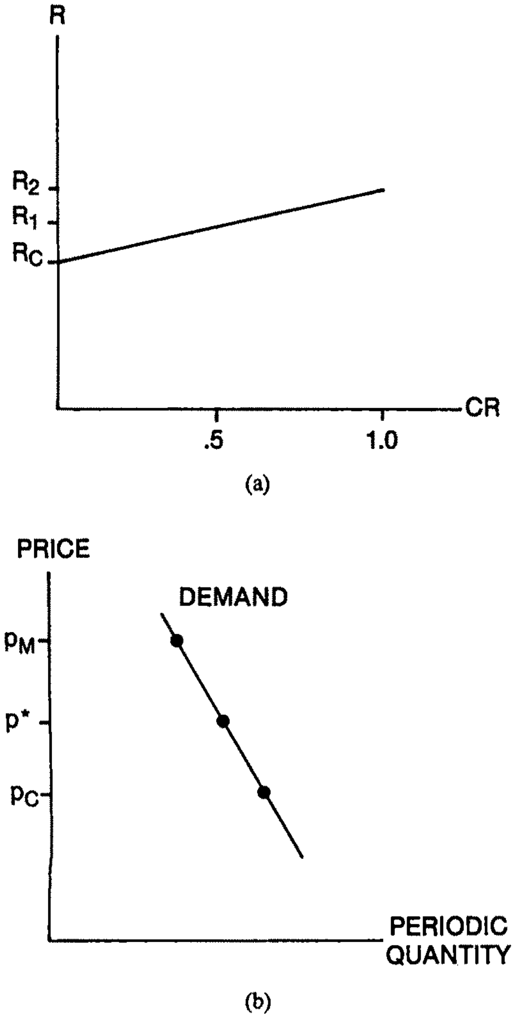


FIGURE 1

not "noncompetitive." They imply prices that are close to the equivalent of p_C, not prices that range between the equivalents of p_M and p^*.

To develop this argument, an equation is derived that, when coupled with available price elasticity data, makes it possible to specify what the magnitude of R would be if typical leading firms *did* price at the equivalent of p_M. Then it is shown that rates of return between the equivalent of R_1 and R_2 are small fractions of this profit-maximizing R and thus are "low," not high. Put another way, it is shown that if, as statistical studies suggest, the price elasticity of demand, η, in most industrial product markets does not exceed 1.5 or 2.0, then there probably are few oligopolists who earn rates of return over time which imply they are maximizing periodic profits and thus generating ratios of pre-tax profits to sales that equal η^{-1}. Instead, most oligopolists appear to earn rates of return which imply ratios of pre-tax profits to sales that are much closer to zero than to η^{-1}.

I

I begin by focusing first on a particular oligopolistic market, and then on one of the leading firms in this market. It is assumed that pricing in the market is governed by a price leadership model, that our firm currently acts as leader, and that all incumbents are about equally efficient. The task is to answer this question: if the firm in its role as leader *were* to choose the price-quantity combination that maximizes its own periodic, after-tax profits, π^τ (and thus π^τ for each of its rivals), what rate of return to invested equity capital, R, would it earn?

First, let us define R and π^τ for any period (say, a year) as

$$R = E/K_e, \tag{1}$$

$$\pi^\tau = E - K_e k_e, \tag{2}$$

where E designates earnings or net income (the difference between sales and the sum of accounting costs, C, interest payments, F, and taxes), K_e designates beginning-of-the-period invested, or book, equity capital (the difference between total invested capital and invested debt capital),[7] and k_e designates the percentage cost of invested equity capital (the equilibrium expected rate of return on

[7]If the firm were newly organized, its total invested capital, K, would equal the reproduction value of the assets in which the capital is embodied. But since the firm is established, K is the divestment value (the opportunity cost of foregoing exit) of its assets. Note that *over time* E and K_e *both* are impacted (and in the same direction) by the way the firm accounts for depreciation, research, and advertising.

each share of the firm's common stock).[8] In addition, define the firm's tax-adjusted average total costs, ATC^{τ}, in the period as

$$ATC^{\tau} = \{C[1-\tau]+F[1-\tau]+K_e k_e\} \div q, \tag{3}$$

where τ is the tax rate and q is quantity produced and sold.

Next a construct is introduced denoted α and defined as the ratio of $K_e k_e$ to $[ATC^{\tau}]q$. This construct is a positive function of the degree of the firm's capital intensiveness and a negative function of the percentage of total invested capital the firm finances with debt.

Finally, let us (i) write the definition of earnings to read

$$E = pq[1-\tau]-[ATC^{\tau}]q+K_e k_e; \tag{4}$$

(ii) assume that over the relevant range of quantities long-run ATC^{τ} approximates tax-adjusted, long-run marginal costs, MC^{τ}; (iii) specify that when π^{τ} is maximized,

$$MC^{\tau} = p[1-1/\eta][1-\tau]; \tag{5}$$

and (iv) make appropriate substitutions (based on equation (5), the assumed relationship between ATC^{τ} and MC^{τ}, and the definition of α) for E and K_e in equation (1) and simplify. We get

$$R = k_e\{1+\alpha[\eta-1]\}/\{\alpha[\eta-1]\}. \tag{6}$$

This equation tells us that if our firm prices at the equivalent of p_M in Figure 1b, the magnitude of R relative to k_e will be an *inverse* function of the magnitude of η at p_M, given α and τ.[9] Table 1 shows what this function looks like, first assuming that k_e equals 11 percent and α equals either 5 or 10 percent, and then assuming that k_e equals 14 percent and α equals either 10 or 15 percent. A presumption here is that for most industrial firms α lies somewhere between 5 and 15 percent when k_e is somewhere between 11 and 15 percent. This presumption is based on Federal Trade Commission data.[10]

TABLE 1—THE MAGNITUDE OF THE FIRM'S PROFIT-MAXIMIZING R FOR VARIOUS MAGNITUDES OF η, k_e, AND α[a]

	$k_e=11$		$k_e=14$	
η	$\alpha=5$	$\alpha=10$	$\alpha=10$	$\alpha=15$
50	16	13	17	16
10	35	23	30	24
5	66	39	49	37
4	121	66	84	61
2	231	121	154	107
1.5	451	231	294	201
1.2	1,111	561	714	481

[a] R, k_e, and α are expressed as percents.

For most goods produced in the industrial sector η appears to run below 1.5.[11] Further, under "cooperative" pricing in general and the price leadership model in particular, η at p_M for each leading incumbent's demand schedule should equal η at p_M for the market schedule. Equation (6) implies directly, then, that if our market is typical and if our firm prices to maximize π^{τ} each period, it will generate a "spread" between R and k_e of 200 percentage points or more (and much more if η is closer to 1.0 than to 1.5). At this price the ratio of π^{τ} to sales equals $\eta^{-1}[1-\tau]$.

Equation (6) also implies indirectly that if our market is typical and if our firm prices to generate an $R-k_e$ spread somewhere between 10 percent and zero, there is a pre-

[8] For investors with an n-year horizon k_e is the discount rate which in equilibrium equates the current price of each share of the firm's stock and the present value of the dividend stream and liquidating price they forecast the share will generate.

[9] Since k_e varies directly and α varies inversely with the percentage of total capital financed with debt, k_e and α are themselves linked. Assuming that the financial leverage model developed by Franco Modigliani and Merton Miller (1963) holds, a significant increase in the ratio of debt capital to total capital increases k_e only modestly.

[10] The FTC presents in each issue of its *Quarterly Financial Report for Manufacturing, Mining and Trade Corporations* estimates of the structure of recent financial statements of samples of corporations operating in various industrial sectors. The estimates are in a form which makes it possible to infer the magnitude of α for each sample for any year, given assumptions about k_e and τ.

[11] There is a useful summary of existing empirical knowledge about price elasticities in Worcester (1975).

sumption the firm is limiting the ratio of π^{τ} to sales to a magnitude much closer to zero than to $\eta^{-1}[1-\tau]$. How much closer? Simulation experiments (using a variety of plausible pairs of linear demand and constant, tax-adjusted average total cost schedules)[12] suggest that the ratio of π^{τ} to sales would be roughly one-half the $R-k_e$ spread.[13] For example, if for our price leader it were the case that $p=40-.008q$ and ATC^{τ} equals 4.005,[14] then we would get these relationships ($R-k_e$ and π^{τ}/pq are shown in percent):

$\frac{p[1-\tau]}{ATC^{\tau}}$	$R-k_e$	$\frac{\pi^{\tau}}{pq}$
3.00	217	33.3
2.00	108	25.0
1.09	10	4.2
1.04	7	3.0
1.02	2	0.9
1.01	1	0.4
1.00	0	0

At the tax-adjusted price-cost ratio of 3.00, π^{τ} is maximized, η equals 1.5, and α equals 12 percent.[15]

The profit-sales ratio is also about one-half the size of a "small" $R-k_e$ spread if the firm faces a constant-elasticity demand schedule and/or a bowl-shaped ATC^{τ} schedule.

II

What kind of $R-k_e$ spreads *do* leading firms in the industrial sector typically earn: π^{τ} maximizing, in the neighborhood of zero, or somewhere in between? Unfortunately, for two reasons this question cannot be answered directly. First, many of the largest producers in this sector are business units of multiproduct, multimarket corporations, and none of these corporations makes public the rates of return their units earn.[16] Second, it is impossible in all but a small minority of 5-digit and 7-digit product markets to find out who the leading firms in fact are.

We can answer the question indirectly, however, by: (a) assuming that Proposition 1 (restated to read that $R-k_e$ *spreads* increase with concentration) is approximately factual across all industrial markets, appropriately defined; (b) identifying a large sample of industrial corporations; (c) estimating the average annual $R-k_e$ spreads generated by this sample over an appropriate study interval;[17] and (d) interpreting each of these spreads as an approximation of the spread generated by each of the corporation's business units.

My sample is the 777 U.S. corporations operating in the manufacturing and trade sectors which are listed by *Forbes* in its 30th "Annual Report on American Industry," 1978.[18] The study interval is the five years 1973–77.

Table 2 shows the distribution of average annual rates of return to invested equity capital earned over this five-year interval by these 777 corporations. The median (and approximate average) of the distribution is around 13 or 14 percent. Over the same interval the Standard & Poor "400 Industrials," the Value Line "Industrial Composite," and the FTC samples of manufacturing corporations all generated an average rate of

[12]Tax adjusted average total costs are defined, again, by (3).

[13]These pairs of experimental demand and cost equations all were constructed so that they would imply at the π^{τ}-maximizing quantity: (a) an η somewhere between 1.2 and 1.7; (b) an α somewhere between 8 and 15 percent; and (c) a ratio of debt capital to total capital somewhere between 30 and 40 percent.

[14]This particular schedule results from the assumptions that τ equals 0.5, k_e equals 13 percent, C equals $6.90q$, K_e equals $3.69q$, total capital equals $6.15q$, and the cost of debt capital equals 7 percent.

[15]At the price-cost ratio of 3.00, the earnings, or income, "margin" equals 35 percent; at the price-cost ratio of 1.02, this margin equals 7 percent. The earnings margin is the ratio of E to sales. It reflects not only the ratio of π^{τ} to sales, but also the leverage ratio and the ratio of sales to invested capital.

[16]Scherer (pp. 74–78) summarizes what is currently known about the extent to which large corporations have diversified themselves. The "10K" reports now being published by multi-business companies only display sales, income, and asset data for *groupings* of individual business units.

[17]An "appropriate" study interval is one which minimizes the likelihood that spread estimates will be biased significantly either by the way events in some years are accounted for or by the impact of unusual demand-side or cost-side disturbances.

[18]I assume that this sample exhibits the same α range that the FTC samples (discussed in fn. 10) do.

TABLE 2—THE DISTRIBUTION OF RATES OF RETURN TO INVESTED EQUITY CAPITAL EARNED ON AVERAGE BY 777 *FORBES* INDUSTRIAL CORPORATIONS OVER THE INTERVAL 1973–77

Average Rate of Return to Invested Equity Capital[a]	Percentage of Corporations
Greater than 30.0	2
25.0–29.9	4
20.0–24.9	11
17.5–19.9	10
15.0–17.4	17
12.5–14.9	19
10.0–12.4	15
7.5–9.9	10
5.0–7.4	5
2.5–4.9	2
Less than 2.5 including negative	7

Source: Forbes, "30th Annual Report on American Industry," January 2, 1978.
[a]Shown in percent.

return of around 14 percent.[19] About one-fifth of the *Forbes* companies had rates of return in the 12.5–14.9 percent cell containing the median. About two-fifths of the companies had rates of return less than 12.5 percent, about two-fifths had rates of return greater than 15 percent, and about one-twentieth had rates of return between 25 and 35 percent.

For any corporation the cost of equity capital can be expressed $k_e = k + r$, where k is the nominal rate of return on long-term U.S. Treasury bonds and r is the risk premium asked by owners. Thus for any sample of corporations the *average* k_e can be expressed $k_e = k + r_A$, and the k_e for each individual corporation in the sample can be expressed

$$k_e = k + [r_A]\beta, \qquad \beta \gtreqless 1,$$

where r_A is the average risk premium for the sample and β can be thought of as a "risk multiplier." Estimates of risk premia are difficult to make, but for the study interval and for every available sample of large industrial corporations (including the *Forbes* sample), there seems to be a rough consensus among investigators that r_A was somewhere around 5.5 percent and that the risk multipliers of a large majority of the sample members were distributed approximately normally between 0.7 to 0.8 and 1.3 to 1.4.[20] Since k averaged about 7.5 percent over the interval, it follows that most of the corporations in the *Forbes* sample had a k_e somewhere between 11 or 12 percent and 15 or 16 percent, and that the average k_e for the sample equalled about 13 percent. There are no estimates which indicate that corporations at the top of the distribution in Table 2 typically had costs of equity capital significantly lower than the sample average.

It appears then that, on average, the 777 corporations in the *Forbes* sample generated spreads between R and k_e which were not much different from zero, that about one-third of the sample generated spreads which were clearly positive. About one-sixth of the sample appears to have generated spreads which were greater than 5 percent and about one-twentieth of the sample appears to have generated spreads which were between 10 and 20 percent.

This same pattern also appears to have prevailed in earlier study intervals. In particular, *Forbes* (1973, 1968) and FTC data, combined with reasonable assumptions about the behavior of risk premia since 1960,[21] indicate that for large U.S. industrial corporations the mean and distribution of $R - k_e$

[19]The Standard & Poor and Value Line rate of return estimates come from periodic reports published by these organizations. The FTC estimate comes from various issues of the *Quarterly Financial Report*....

[20]Most investigations of corporate capital costs have been carried out by investment specialists. Wells Fargo Bank, for example, has estimated regularly for about a decade the expected, or required, rate of return (i.e., cost of equity capital) on the common stock of each of over 300 large industrials (almost all of which are included in the *Forbes* sample). Today every major investment house (including Merrill Lynch and Value Line) directly or indirectly estimates costs of equity capital for large industrials (over 900 in the case of Merrill Lynch). For a textbook exposition of the meaning and measurement of the cost of equity capital (and for estimates of r_A), see Lawrence Schall and Charles Haley (1983, ch. 7).

[21]These assumptions are that r_A and the shape of the distribution of risk multipliers both have been roughly stable since 1960. George Racette (1972) offers some support for the first assumption.

spreads were about the same over the intervals 1963–67 and 1968–72 as they were over the interval 1973–77.[22,23]

An obvious question intrudes itself. Is there not a possibility that some corporations in the sample consistently generated consolidated $R-k_e$ spreads of around zero because half of their business units generated consistently high positive spreads (say, around 100 percent) while the other half generated consistently low negative spreads (say, around -100 percent)? There is, but it seems remote.

First, consider this distribution of the 21 "industries" into which *Forbes* has grouped the 777 corporations in Table 2. ("Equal Risk" shown in percent):

Number of Industry Groups	"Equal Risk" $R-k_e$ Spreads
1	6
5	3 or 4
6	1 or 2
2	0
6	-1 or -2
1	-3

The equal risk $R-k_e$ spread for each industry group is the difference between the average R for the group and 13 percent, the average k_e for the whole sample. Admittedly, these groups probably are *not* equally risky, and thus can be presumed to have had *different* rather than identical costs of equity capital. But published estimates of capital costs for other industry groupings suggest first, that most costs lie within a band about five percentage points wide,[24] and, second, that those groupings which have the lowest capital costs do not tend to earn the highest rates of return.[25] Accordingly, it seems safe to interpret this distribution as at least a rough approximation of the distribution of actual spreads for the *Forbes* sample.[26] It also seems safe, then, to believe the message the distribution sends: few if any business units in diversified corporations would have been *able* to generate large positive $R-k_e$ spreads.[27]

Second, not *all* producers in oligopolistic industries are units of multibusiness corporations. Some are, in effect, single industry firms, or if not, confine their operations to closely related 5-digit or 7-digit product markets. And there is no evidence at all that a significant number of *these* producers consistently generate spreads of more than 4 or 5 percent.

Summing up, then: further empirical work possibly could show that Proposition 1 restated is counterfactual. But even if, as we are assuming here, the proposition does hold at least approximately, our spread estimates tell us that Proposition 2 restated (higher spreads also are high) does not.[28] There sim-

[22]This reading of the 15-year interval implies of course that in the aggregate the sample corporations increased R from 1963 to 1977 by about the same number of percentage points their average k_e increased. The latter increased in turn by about the same number of percentage points the rate of inflation increased.

[23]In each of the 5-year intervals, somewhere between one-third and one-half of large industrials apparently generated negative $R-k_e$ spreads. Why? What explains the unwillingness of corporations to divest their "unprofitable" units? Some (especially capital intensive ones) have concluded that while their negative spread units are worth less than book value, they are worth *more* than divestment value. (If the book values of these units were written down to going concern values, of course, spreads would be zero.) Other corporations have looked into the future and seen negative spreads being replaced with large, positive ones. And still others have chosen to measure performance, not by π^τ but by income; for them exit is mandated only by red ink.

[24]This statement is based on estimates made by Wells Fargo Bank, Merrill Lynch (1981), and Barr Rosenberg and James Guy (1976).

[25]For example, the *Forbes* industry group "Consumer Goods: Personal Products" generated the highest R over the study interval. Estimates by Merrill Lynch imply that the risk multiplier for this group equals about 0.9; estimates by Rosenberg and Guy imply that the multiplier for this group equals about 1.1 or 1.2.

[26]The corporations represented at the top of the distribution in Table 2 are drawn from all 21 industry groups.

[27]There are no available concentration data for the *Forbes* groups (because they do not match the SIC 2-digit groups). Judging from the Proposition 1 studies, though, there presumably *is* some sort of modestly positive, but noisy, relationship between average spreads and average concentration ratios for the *Forbes* groups.

[28]Since k averaged around 6 percent from 1968 to 1973, the average k_e for large industrial corporations probably averaged no less than 11 percent over this interval. Thus under Title 1 of the Hart Bill (see fn. 4), a firm of average riskiness which generated an average $R-k_e$ spread of 4 percent over the interval 1968–72

ply is no support in the data for the contention that leading firms in concentrated industries cooperatively choose locations on their rate of return schedules that are significantly above the equivalent of R_C and on their share-of-the-market demand schedules that are significantly above the equivalent of p_C.[29,30] Noncompetitive structures do not result in noncompetitive performance.[31]

III

Proposition 2 restated is really a surrogate for this joint hypothesis about leading firms in concentrated industries:

(i) A two-part model guides their thinking: (a) because of the cost barriers which surround them, the magnitude of profits in years $1,2,3,4,\ldots$ will have little impact on the rate of entry in these years; and (b) in the light of this independence in the profits vector, the way to maximize the present value of this vector,

$$\pi_1^{\tau}/(1+k_e)+\pi_2^{\tau}/(1+k_e)^2+\pi_3^{\tau}/(1+k_e)^3 + \pi_4^{\tau}/(1+k_e)^4+\ldots,$$

is to maximize, or at least come close to maximizing, first π_1^{τ}, then π_2^{τ}, then π_3^{τ}, and so forth.

(ii) They succeeded in implementing this model.

Thus, if the surrogate probably should be abandoned, so should the joint hypothesis. What should replace it? How can the performance we observe in the industrial sector be explained? There are several possibilities.

One is that leading firms try to price cooperatively at levels which will maximize in sequence $\pi_1^{\tau}, \pi_2^{\tau}, \pi_3^{\tau},\ldots,$ but are thwarted by follower firms who as a means of increasing their market shares cut price to the neighborhood of average total costs. The leaders, faced with a choice between maintaining price and maintaining their own shares, decide that the present value, V, of expected annual profits will be greater if they match the price cuts than if they do not. In short, cooperative pricing breaks down in all industries.

A second hypothesis is that leading firms perceive barriers to be so low that they can deter entry only by constraining $R-k_e$ spreads to very modest levels. Faced with a choice between cooperatively pricing to generate these spread levels and cooperatively pricing to generate higher ones, the leaders decide that V will be higher if they maintain share (or keep share from falling more than a few points) than if they don't. In short, barriers to entry are not high enough anywhere to make noncompetitive spreads consistent with value maximization.

A third hypothesis is that some *combination* of uncooperative pricing and modest barriers works to keep essentially competitive ceilings on profitability.

Time will tell which of these hypotheses best explains this study's findings. There is

---

could be presumed to possess unlawful monopoly power. My simulation experiments suggest, again, that a spread of 4 percent implies a ratio of π^{τ} to sales of no more than 2 percent.

[29] Gary Becker (1971, p. 97) appears to argue: (a) that accountants for leading oligopolists write up book values to market values; and (b) that the rates of return of these business units therefore necessarily will approximate their costs of equity capital. In fact, while business unit accountants can (and sometimes do) write down book values, they cannot (and do not) write them up.

[30] George Stigler (1956) has raised the possibility that some corporations in the past may have acquired business units at prices which capture the fruits of π^{τ} maximizing conduct, with the result that the subsequent rates of return of these corporations give a misleading picture of the subsequent conduct of the acquired units. In fact, acquisition prices recorded over the last two decades or so imply that acquired business units typically were expected to earn essentially competitive rates of return as independent firms. Further, when a corporation finances an acquisition with *common shares*, it records on its own balance sheet, not the price of the new unit, but the unit's book value.

[31] J. B. Clark (1907), P. W. S. Andrews (1949), R. F. Harrod (1952), H. R. Edwards (1955), and others hypothesized long ago, in effect, that the data would show this. In more recent times, Yale Brozen (1970) can be interpreted as having hypothesized the same thing.

no need, however, to defer drawing the appropriate policy conclusions from the findings.

REFERENCES

Andrews, P. W. S., *Manufacturing Business*, London: Macmillan, 1949.

Bain, Joe S., *Barriers To New Competition*, Cambridge: Harvard University Press, 1956.

Becker, Gary S., *Economic Theory*, New York: Alfred A. Knopf, 1971.

Brozen, Yale, "The Antitrust Task Force Deconcentration Recommendation," *Journal of Law and Economics*, October 1970, *13*, 279–92.

Buzzell, Robert D., "Are There 'Natural' Market Structures?," *Journal of Marketing*, Winter 1981, *45*, 42–51.

Clark, John B., *The Essentials of Economic Theory*, New York: Macmillan, 1907.

Demsetz, Harold, "Industry Structure, Market Rivalry, and Public Policy," *Journal of Law and Economics*, April 1973, *16*, 1–10.

______, "Two Systems of Belief About Monopoly," in Harvey J., Goldschmid et al., eds., *Industrial Concentration: The New Learning*, Boston: Little Brown, 1974.

Edwards, H. R., "Price Formation in Manufacturing Industry and Excess Capacity," *Oxford Economic Papers*, February 1955, *7*, 94–118.

Goldschmid, Harvey, J. Mann, H. Michael and Weston, J. Fred, *Industrial Concentration: The New Learning*, Boston: Little Brown, 1974.

Hay, Donald A. and Morris, Derek J., *Industrial Economics, Theory and Evidence*, Oxford: Oxford University Press, 1979.

Harrod, R. F., *Economic Essays*, London: Macmillan, 1952.

Koch, James V., *Industrial Organization and Prices*, 2nd ed., Englewood Cliffs: Prentice-Hall, 1980.

Leftwich, Richard, *The Price System and Resource Allocation*, 6th ed., Hinsdale: Dryden, 1976.

Mansfield, Edwin, *Microeconomics*, 4th Edition, New York: Norton, 1982.

Modigliani, Franco and Miller, Merton H., "Corporate Income Taxes and the Cost of Capital," *American Economic Review*, June 1963, *53*, 433–43.

Mueller, Dennis C., "The Persistence of Profits Above the Norm," *Economica*, November 1977, *44*, 369–80.

Peltzman, Sam, "The Gains and Losses from Industrial Concentration," *Journal of Law and Economics*, October 1977, *20*, 229–63.

Racette, George, "Risk and the Required Rate of Return," unpublished doctoral dissertation, University of Washington, 1972.

Rosenberg, Barr and Guy, James, "Prediction of Beta From Investment Fundamentals," *Financial Analysts Journal*, July-August 1976, *32*, 62–70.

Samuelson, Paul A., *Economics*, 11th ed., New York: McGraw-Hill, 1981.

Schall, Lawrence D. and Haley, Charles W., *Introduction to Financial Management*, 3rd ed., New York: McGraw-Hill, 1983.

Scherer, F. M., *Industrial Market Structure and Economic Performance*, 2d ed., Chicago: Rand McNally, 1980.

Sherman, Roger, *The Economics of Industry*, Boston: Little Brown, 1974.

Shepherd, William G., *The Economics of Industrial Organization*, Englewood Cliffs: Prentice-Hall, 1979.

Stigler, George J., "The Statistics of Monopoly and Merger," *Journal of Political Economy*, February 1956, *64*, 33–40.

Weiss, Leonard, "Quantitative Studies of Industrial Organization," in Michael D. Intriligator, ed., *Frontiers of Quantitative Economics*, Amsterdam: North-Holland, 1971.

______, "The Concentration-Profits Relationship and Antitrust," in Harvey J. Goldschmid, et al., eds., *Industrial Concentration: The New Learning*, Boston: Little Brown, 1974.

Worcester, D. A., Jr., *Monopoly, Big Business and Welfare in the United States*, Seattle: University of Washington Press, 1967.

______, "On Monopoly Welfare Losses: Comment," *American Economic Review*, December 1975, *65*, 1015–23.

Federal Trade Commission, *Quarterly Financial Report for Manufacturing, Mining, and*

Trade Corporations, Washington, various years.

Forbes, "20th Annual Report on American Industry," January 1, 1968; "25th Annual Report on American Industry," January 2, 1973; "30th Annual Report on American Industry," January 2, 1978.

Merrill Lynch, *Quantitative Analysis, Common Stock Valuation, A Bimonthly Report*, November 1981.

Standard & Poor Corporation, *Standard & Poor's Analysts Handbook, Official Series*, 1982 Annual Edition.

Value Line, *Investment Survey*.

Wells Fargo Investment Advisors, Wells Fargo Bank, *Security Market Line*.

Some Calculations of Lifetime Tax Incidence

By JAMES DAVIES, FRANCE ST-HILAIRE, AND JOHN WHALLEY*

This paper reports a set of lifetime tax incidence calculations using a life cycle simulation model for Canada due to Davies (1979a, 1982). A repeatedly stated qualification to annual calculations in the empirical tax incidence literature is that it would be more satisfactory to make calculations on a lifetime basis. Even though it is acknowledged that lifetime tax incidence could well differ from annual, it is widely believed that data and other difficulties make such calculations next to impossible. Indeed, the widespread acceptance of the data problems of lifetime calculations seems also to have inhibited speculation about how lifetime tax incidence might differ from annual. As a result, redistributive tax policy judgments continue to be based on annual incidence calculations in spite of the reservations many have about their usefulness. Our paper is intended to reorient discussion towards lifetime tax incidence by providing some initial null hypotheses about the shape of lifetime tax profiles.

Our main finding is that under the standard competitive assumptions common in the incidence literature, lifetime and annual incidence calculations both produce mild progression in tax rates across household deciles (ignoring the bottom decile in the annual calculation). While the income tax is less progressive in lifetime than in annual calculations, other taxes are for the most part less regressive. Also, lifetime incidence calculations are much more robust to alternative shifting assumptions than annual calculations. In the lifetime context, key distributions such as earnings, transfer payments, and consumption are less heavily concentrated in particular percentiles of the population than is true in annual data. As a result, changing the allocative series for any particular tax does not have the large effect on incidence results found in annual calculations.[1]

Each component of the tax system is allocated to households grouped by lifetime income using particular distributive series following a procedure similar to that employed in annual incidence calculations (for example, Richard Musgrave et al., 1974; Joseph Pechman and Benjamin Okner, 1974; Edgar Browning and William Johnson, 1979; W. Irwin Gillespie, 1980). In the process we are able to compare lifetime and annual incidence calculations using the same data set.

In both lifetime and annual calculations, we allocate five groups of taxes among households using distributive series which come partly from the 1971 Statistics Canada Survey of Consumer Finances (*SCF*) and partly from our life cycle simulation model. The *SCF* data are used to construct synthetic longitudinal lifetime profiles of earnings and transfer payments for a sample of 500 households. The latter are assigned inheritances by simulating patterns of mortality and bequest. These data are then used in the life cycle model to generate lifetime consumption profiles and bequests. The earnings, transfer, and inheritance data, plus the model output provide the distributive series on which alternative incidence calculations are based.

While the incidence calculations presented in this paper use Canadian data, results would likely be similar for the United States

*Department of Economics, University of Western Ontario, London, Canada N6A 5C2. We thank participants in workshops at McMaster University and the University of Western Ontario, as well as an anonymous referee, for useful comments. St-Hilaire and Whalley gratefully acknowledge financial support from the Social Sciences and Humanities Research Council of Canada.

[1]See St-Hilaire and Whalley (1982) for an investigation of the impact of changing incidence assumptions in annual tax incidence calculations. We note that several of the issues raised by St-Hilaire and Whalley, including the implications of the open economy and the treatment of human capital for tax incidence calculations, are not taken up in this paper. Lifetime calculations may well be sensitive to the way these are treated, and this should be explored in future work.

as is true of annual calculations (see St-Hilaire and Whalley, 1982). Relative to overall tax revenue, personal and corporate income taxes are of roughly the same importance in the two countries. The personal income tax in Canada has a flatter rate profile, less significant exemptions and deductions, and more extensive tax shelters than its U.S. counterpart (for example, mortgage interest is not deductible, but there are larger shelters for savings for retirement). Taxes levied by the provinces are relatively more important than those at state level in the United States. Finally, sales and excise taxes are higher than in the United States, but social security taxes are considerably lower. This latter feature produces an offsetting effect in standard incidence calculations where both taxes are largely regressive.

I. Annual Incidence Calculations and Lifetime Issues

A. *Annual Incidence Studies*

Tax incidence calculations which use annual data typically focus on five key taxes or tax groups: income, corporate, sales and excise, property, and social security. Each tax is treated as having sources (income) side and/or uses (expenditure) side effects for each group of households who, in turn, are usually stratified by income. Incidence calculations for whole tax systems use a separate tax burden calculation for each tax by income range. The overall redistributive effect of the tax system is usually evaluated by examining the pattern of combined average tax rates across income ranges.

Three main income sources—capital, labor, and transfer income—can bear the burden of taxes either singly or jointly. In annual data, transfers are heavily concentrated in the lower tail of the income distribution. Capital income is a relatively high proportion of total income in the upper tail, but also in the lower tail due to the presence of retirees. Labor income increases sharply as a fraction of total income for the bottom few deciles, but is approximately proportional to income for about 70 percent of the population. Thus, on the sources side, widely varying patterns of progressivity or regressivity can result depending upon whether a tax is allocated to capital, labor, or transfer income, or to income in general.

On the uses side, the strongest effects occur through differences in household saving rates. This is because a tax assumed to be forward shifted in higher prices is treated as borne by households out of consumption. In annual data, saving rates differ sharply by income range, with around 70 to 80 percent of household saving commonly concentrated in the top 10 percent of the income distribution.[2] Forward shifted taxes allocated to consumption are therefore regressive.

One of the most widely cited annual incidence studies is Pechman and Okner, which uses detailed merged data involving approximately 87,000 1966 U.S. income tax returns and 30,000 households appearing in a Survey of Economic Opportunity data file. Alternative incidence calculations are made using different shifting assumptions as to the burden of certain taxes. Their main conclusion is that "regardless of the incidence assumptions, the tax system is virtually proportional for the vast majority of families in the United States" (p. 64).[3] This, however, omits the bottom 10 and top 3 percent of the population (see Pechman and Okner, pp.

[2]St-Hilaire and Whalley (1983, Table 9, p. 195) indicates saving of the top decile of Canadian households in 1972 made up 73 percent of aggregate saving. Browning and Johnson (Table 18, p. 73) have a saving rate of 34.4 percent of disposable income for their top decile of 1974 U.S. households. From their Table 9 on page 56 it appears that the top decile would have had about 30 percent of aggregate disposable income. This implies a share of about 76 percent of aggregate saving.

[3]Musgrave et al. also report rough proportionality using assumptions close to Pechman and Okner's least progressive variant and a different data set. Gillespie (1980) obtains comparable results for Canada, and performs sensitivity analysis similar to that of Pechman-Okner. The similarity of results to the U.S. studies is obscured by the use of an income concept which excludes transfer payments. Excluding transfers significantly increases tax rates for bottom income groups, and tends to make the overall tax system appear regressive. When Gillespie's results are presented using an income base comparable to that of Pechman-Okner or Musgrave et al., the similarity in results is more evident. See either St-Hilaire and Whalley (1982, Table 1), or Gillespie (1976, Table A-7, p. 445).

5–6), which has an effect on the conclusions. In their most progressive variant, for example, where corporate and property taxes are borne by capital, the total effective tax rate for the bottom 10 percent is much lower than for other deciles and that for the top 3 percent far above average (see Tables 4.4 and 4.9 on pp. 51 and 61, respectively).

This "proportionality" view of overall tax incidence has been questioned in recent work by Browning (1978) and Browning and Johnson (1979). The main difference concerns the treatment of sales and excise taxes which are regressive in Pechman-Okner and progressive in Browning-Johnson. Browning and Johnson argue that uses side effects of sales and excise taxes due to different saving rates by income range are insignificant when consumption out of normal or permanent income is considered. They therefore only consider the sources side effects of sales and excise taxes, pointing out that since transfers are largely indexed for price level changes, only factor incomes can bear the burden of taxes. The concentration of transfers in the lower tail of the income distribution, and saving in the upper tail means that sales and excise taxes which appear as strongly regressive in Pechman-Okner are strongly progressive in Browning-Johnson.[4]

The incidence assumptions used in this literature are crucial to the extent of progressivity in results. The income tax is universally treated as paid by income recipients and is progressive. However, widely varying assumptions for other taxes are examined in the literature. Social security is treated either exclusively or predominantly as a payroll tax on labor and (outside of the lower tail of the income distribution) is regressive due to ceilings on contributions. Corporate and property taxes are regressive if treated as shifted forward onto consumption, but mainly progressive if assumed shifted backwards onto capital income.[5] Finally, sales and excise taxes are regressive if borne by consumers, and progressive if treated as borne by recipients of factor incomes. Depending on the combination of shifting assumptions the tax system can be made to appear regressive or progressive in annual calculations. (See St-Hilaire and Whalley, 1982.)

B. *Lifetime Incidence — Expected Contrasts with Annual Calculations*

Despite the extensive literature reporting annual incidence calculations, it is widely acknowledged that lifetime calculations would be preferable.[6] Data limitations are usually cited as precluding such an approach. (See, for example, Browning and Johnson, p. 81.) The lack of incidence calculations on a lifetime basis is especially unfortunate, since on a priori grounds there are several reasons to believe there may be significant differences between annual and lifetime incidence results. First, much of the observed inequality in earnings and transfer payments disappears when we examine lifetime rather than annual distributions. According to the estimates of Jacob Mincer (1974), Lee Lillard (1977), and Nils Blomqvist (1981), about one-half of annual earnings inequality (according to conventional measures) disappears when one looks at lifetime earnings.[7] Social security and other in-

[4]An intermediate case would ignore the indexation of transfers on the sources side but include consumption out of permanent income on the uses side. Sales and excise taxes may then have uses side effects arising from differing taxes on different consumer goods. See David Davies (1959) for an early study of sales and excise tax incidence that proceeds along these lines.

[5]Corporate taxes are more progressive if treated as borne by capital specific to taxed industries rather than capital in general, because of light tax treatment of widely held housing capital. In some of the literature, this observation motivates the use of dividends rather than capital income in general as a more progressive distributive series for allocating corporate taxes.

[6]See, for example, Browning and Johnson (pp. 24–26 and 80–84) and Anthony Atkinson and Joseph Stiglitz (1980, p. 286). Acknowledgement that lifetime calculations would be preferable is also implicit in the repeated argument that the uses side effects of sales and excise taxes should be evaluated on the basis of patterns of consumption out of permanent income. See, for example, D. Davies, and Pechman-Okner (pp. 52–55).

[7]Mincer (p. 119) argues that about half the log variance of earnings is attributable to transitory components and age-related differences. Lillard's panel evidence indicates a Gini coefficient for lifetime earnings about 25–40 percent below that in age-specific annual earnings (see p. 50). The short-fall below the annual Gini for all

come support payments, in particular, are much more equally distributed when examined on a lifetime rather than an annual basis. Lifetime transfers are therefore not concentrated chiefly in the hands of the poor as is true in annual data. These differences imply less progressivity of personal income taxes in the lifetime context. Also, with transfers less important and earnings more important in the bottom tail the significance of the indexation issue stressed by Browning and Johnson is weakened.

A second issue relevant to lifetime calculations is the treatment of capital income. This important element in annual incidence calculations is not part of a household's discounted lifetime income. It therefore cannot bear the burden of taxes on the sources side in the same way as in annual calculations since it is not part of the income base. As a result, a fundamentally different treatment of taxes thought to be borne by capital (for example, corporate, property) is required.

A third important difference between annual and lifetime calculations arises from the differing time profiles of taxes over the life cycle. Taxes on labor income, for example, are paid until retirement, but those on consumption are paid until death. Hence the relative lifetime importance of various taxes may differ significantly from their relative annual importance, due to the effect of discounting.

Finally, the profile of consumption as a proportion of income will differ considerably between annual and lifetime data. While there is reason to expect higher income groups to save a larger proportion of their lifetime resources for bequests, the lifetime profile of savings rates by income range ought to be much flatter than the annual. Therefore sales and excise taxes treated as borne by consumers should appear less regressive in the lifetime calculations. Also, switching the incidence assumption for corporate or property taxes from full burden on capital income to partial forward shifting should produce a less dramatic shift towards regressivity.

ages taken together would be higher. Finally, Blomqvist estimates that the Gini for lifetime earnings is about 45–55 percent below that in age-specific annual earnings (see p. 255).

II. A Life Cycle Approach to Incidence Analysis

Later in the paper we report tax incidence calculations based on lifetime rather than annual data. Our approach is similar to that used in annual calculations in that we adopt alternative distributive series for the allocation of various taxes between household groups. The main differences are that distributive series are generated using lifetime rather than annual data, households are grouped by lifetime income, and taxes paid over the lifetime rather than in a single year are allocated.

Our distributive series rely on the data used in, and the output from, a micro-simulation model of life cycle saving and bequest behavior of Canadian households reported in Davies (1982). In Davies' model, each household includes a husband and wife who start economic life together at age 20 and die together at age 75.[8] Households are assigned realistic (exogenous) streams of earnings, transfer payments, and inheritances over the lifetime, as described below. The simulation generates lifetime paths of consumption and investment income which along with the lifetime income data provide the distributive series for our incidence calculations.

A. *The Behavioral Model*

Each household receives an exogenously determined stream of earnings, E_t, and government transfers, G_t (depending on age, t) over its lifetime. We refer to the sum of earnings and transfers as noninvestment income, N_t. Real interest before tax on wealth, W_t, is received at a constant rate, r, giving rise to investment income, $M_t = rW_t$. Annual

[8]Each married couple has a number of children. Children are taken into account in the determination of bequests and in the simulation of saving behavior. The earnings hump in middle age produces less life cycle saving than would occur in a model without children since up to age 45 most families' consumption requirements are increased by children. The number and spacing of children is designed to be quasi realistic. See Davies (1982) for details.

income gross of (direct) tax, Y_t, is thus given by

$$(1) \qquad Y_t = E_t + G_t + M_t = N_t + M_t.$$

This income is divided between consumption, C_t (gross of indirect taxes); direct taxes, T_t, and saving, S_t.[9] The annual budget constraint is thus:

$$(2) \qquad Y_t = C_t + S_t + T_t.$$

Over the lifetime, resources are augmented by an exogenous stream of capital transfers, I_t, which we refer to as inheritances. Denoting discounted present values over the lifetime by dropping the t subscript, lifetime income or lifetime resources, L, is given by

$$(3) \qquad L = E + G + I = N + I,$$

where discounting occurs at a family's after-tax discount rate, r^*.[10]

Like annual income Y_t, lifetime resources L are gross of direct tax. However, because lifetime investment income M is not included in L, direct taxes on investment income are not included in the lifetime measure of direct taxes on $L(\tilde{T})$ as they are in the annual measure of direct taxes on $Y_t(T_t)$. The $\tilde{T}$ represents direct taxes on E and G (I is net of estate taxes). With this definition, and denoting the discounted value of bequests as B, we have the lifetime budget constraint:

$$(4) \qquad L = C + B + \tilde{T}.$$

Households generate a time path of C_t and a value of B by maximizing an intertemporal utility function defined over consumption and bequests, subject to (4). The utility function is of the familiar additive iso-elastic form (see, for example, Alan Blinder, 1974) but with parents' consumption and children's expected lifetime income (which depends partly on parents' bequests) as arguments. Bequests have a compensatory element since they are directly related to parents' income, but inversely related to children's anticipated earnings. Because the latter regress to the mean, on average, bequests rise as a proportion of parents' income. Lifetime consumption C therefore declines as a proportion of L, although to a much smaller extent than does annual consumption as a fraction of annual income.

B. *The Simulation*

The simulation model assumes that Canada is in balanced growth with labor force growth in efficiency units of 2 percent. Per capita earnings, transfer payments, taxes, and other flows all grow at the same rate, and successive cohorts are identical except for a proportional increase in their resources and expenditures. The balanced growth assumption makes it possible to infer time paths of the distributions of all the annual flows (E_t, G_t, Y_t, etc.) for each cohort from the observed values of a representative cross section at a point in time. The before-tax interest rate r is set at 6 percent.[11]

Lifetime paths of E_t, G_t, T_t, and $\tilde{T}_t$ are generated for each cohort by taking random samples of 500 households for each of 11 different age groups (20–24, 25–29,...,70–74) from the 1971 Statistics Cana-

[9]Direct taxes, T_t, include employees' share of social security premiums (imputed on the basis of reported E_t and the statutory rate schedule), and personal income taxes.

[10]As explained in Section I, Part B, r^*s differ slightly between households as we assume a common before-tax rate of interest, r, and use different marginal tax rates across households. In the tax incidence calculations, where comparisons between households must be made, it is necessary to discount lifetime streams for all families using the same interest rate. This is done using the common rate, r (see Section III, Part B).

[11]An overall before-tax average rate of return of 6 percent for households was used by Blinder in his pioneering study of life cycle behavior and income distribution. Although lower rates are sometimes thought more realistic, we find support for such a rate in some of our earlier work. St-Hilaire and Whalley (1983, Table 10) estimate that aggregate comprehensive household investment income in 1972 was \$16.5 billion. When adjusted to a 1970 basis for growth and inflation, at \$13.0 billion, this is approximately 6 percent of the estimate of aggregate household net worth for 1970 in Davies (1979, Table 2)—\$216.9 billion.

da Survey of Consumer Finances (*SCF*).[12] The data are adjusted to a 1970 basis, correcting for inflation and growth, to make them consistent with the rest of the model data base. Households are ranked according to total reported income in each age group. They are then linked across age groups in order to construct 500 life histories for E_t and G_t. The linking procedure contains a random element, whose strength may be varied to produce more or less earnings mobility. We parameterize this process to reproduce the degree of earnings mobility implicit in Lillard's study of the relationship between annual and lifetime earnings in the United States.[13] Sensitivity of the incidence results to the assumed degree of earnings mobility is checked by repeating our calculations assuming zero and free (random) mobility.

The 1971 *SCF* does not contain any information on inheritances. In order to model this, Davies (1982) produces a cohort distribution of inheritances by simulating the mortality of the Canadian population.[14] Estimates of the size distribution of wealth by age groups for 1970 are available for Canada in the data described by Davies (1979a, b). Assuming that the 1970 data provide a snapshot of a point on a balanced growth path, the age-specific distribution of wealth at any date can be inferred. Simulation of mortality, along with estate division practices and taxes based on observed patterns, produces a size distribution of net inheritances which can be fed into the life cycle saving simulation. This is done in such a way that the correlation between parents' lifetime noninvestment incomes and (assumed) inheritances is close to that between children's income and (simulated) inheritances.

The simulation assigns each household a constant marginal tax rate, u, on investment income over its lifetime.[15] Hence the after-tax rate of interest, $r^* = (1-u)r$. The rate u is calculated by examining each household's reported income tax payment when aged 45–49 in the 1971 *SCF* data. The 1971 tax tables indicate the corresponding marginal tax rate. Since taxable investment income is affected by various exclusions (for example, nontaxation of imputed rent and capital gains on owner-occupied houses, and one-half of all other capital gains), and the lack of accounting for inflation, this legal rate does not correspond to the effective tax rate on true investment income. Legal rates are transformed into effective marginal tax rates as follows. Comparison of aggregate taxable, and simulated total, investment income for 1970 shows that the former was approximately 15 percent of the latter. We make the strong assumption that this ratio is the same for all households. The effective tax rate on broadly defined investment income is therefore set at 15 percent of the legal rate for each household.

III. Implementing Incidence Calculations in the Life Cycle Framework

Our life cycle incidence calculations use the data and the output from the model described above in a procedure commonly used in annual incidence calculations. As a central case we use a set of standard competitive incidence assumptions adapted to the lifetime context and employ alternative

[12]Ideally, $\tilde{T}_t$ would be found by computing the personal income tax (and social security taxes) payable on $N_t = E_t + G_t$ if M_t were zero. (This is appropriate since N_t is exogenous.) Since this computation cannot be made with the *SCF* data (numerous deductions and exclusions are not reported) we have instead prorated the total income tax paid between N_t and M_t. Given the progressivity of the income tax this means that $\tilde{T}_t$ is slightly overstated for each household. This is not a major effect since reported M_t is typically small compared with N_t.

[13]In our central case, the ratio of the Gini coefficient for lifetime earnings to that for age-specific annual earnings is almost precisely the same as found in Lillard. For the age groups 25–34, 35–44, and 45–54, we have ratios of .73, .68, and .60. Lillard obtains ratios of .75, .68, .67, and .61 for groups aged 30–34, 35–39, 40–44, and 45–49, respectively.

[14]The procedure is similar to that employed by Laurence Kotlikoff and Lawrence Summers (1981, Section V).

[15]A varying tax rate would preclude a closed-form solution to the household's consumption plan, and require a more complex simulation model. Given the relatively light overall taxation of capital income, use of constant tax rates on capital income is probably not an important defect in the present exercise.

lifetime distributive series for the five major tax groups listed earlier.

In contrast to the annual calculations where investment income N_t is an important distributive series, and inheritances I_t do not appear, the key distributive series on the sources side in the lifetime calculations are discounted lifetime earnings, E; inheritances, I; and transfers, G. These series sum to the gross of tax lifetime income base, L. On the uses side, two rather than the one key series in the annual calculations (C_t) are used: lifetime investment income M, and lifetime consumption C.

A. *Obtaining Comparable Annual Results*

We not only generate lifetime incidence calculations but also annual estimates. A natural method of generating comparable annual incidence calculations is to use annual cross sections from the life cycle simulation to produce distributive series. However, in the simulation model, households are free from liquidity constraints and dissave rapidly when transitory income is negative. The consumption pattern across annual income deciles is therefore more extreme than in actual annual data. Thus while the simulation provides useful information on patterns of saving over the lifetime, it is not as reliable a guide to patterns of saving on an annual basis.

In order to obtain annual tax incidence results we have used the same primary data source (the 1971 *SCF*) as far as possible to produce the distributive series for allocating taxes.[16] The only series that cannot be obtained in this way is consumption, which we generate by simulation. We have found that when zero earnings mobility is assumed—greatly reducing the importance of transitory income—the simulated annual consumption profile becomes reasonably similar to that reported in other data sources.[17] The annual consumption profile obtained in the zero mobility simulation run was therefore used in the annual incidence calculations reported below.

B. *Lifetime vs. Annual Distributive Series*

Table 1 shows the main features of both the lifetime and annual data used in the incidence calculations reported in the next section.[18] The following contrasts between the annual and lifetime data should be highlighted:

1) There is less inequality in the distribution of lifetime income L, than in that of annual income Y. (Both distributions are gross of direct taxes.) It is therefore not surprising that the lifetime data show a lower variance of personal income tax rates over households ranked by income than the annual data.

[16]However, the investment income concept in these data is different from the broad concept in the life cycle simulation. It excludes imputed rental income on housing and durables, and retained earnings. We have made a rough adjustment to compensate for these exclusions by multiplying up the reported *SCF* investment incomes by the ratio of household broad investment income as reported in St-Hilaire and Whalley (1983) to the aggregate investment income indicated by the 1971 *SCF*. (The St-Hilaire and Whalley data are for 1972, and were adjusted for growth and and inflation to put them on a 1971 basis.) A further, downward adjustment was made to this total so that it would underestimate the National Accounts figure by the same fraction as other forms of *SCF* income. This prevents capital income from taking on an unrealistic significance in these computations.

[17]Using this procedure, consumption declines from 157 percent of income for the bottom decile to 49 percent for the top decile (see Table 1). St-Hilaire and Whalley (1983), in comparison, find that consumption varied from 117 to 64 percent of income from the bottom to top income groups (17 and 9 percent of households, respectively) in the 1972 Canadian data. See their Tables 9 and 10. Others have found considerably higher consumption to income ratios at the bottom end. Browning and Johnson (p. 73), for example, have a consumption to *disposable* income ratio of 149 percent for the bottom decile.

[18]All lifetime magnitudes in this and subsequent tables are discounted using the before-tax rate of return r. We cannot make comparisons across households using the after-tax rate r^* since the latter differs between households. There is little difference between r and r^* on average since the tax rates u on investment income are low—in the range 0–10 percent. Our results are therefore unlikely to be affected by the use of r instead of some average of the r^*s. (Note that to use an average of the r^*s some arbitrary weighting scheme for averaging across households would have to be selected.)

TABLE 1—BASIC DATA IN LIFETIME AND ANNUAL INCIDENCE CALCULATIONS[a]

Decile	Share of Total L	Composition of L: E/L	G/L	I/L	Uses of L: C/L	B/L	$\tilde{T}/L$	M/L
A. Lifetime								
1	4.2	81.2	15.3	3.5	86.7	4.4	8.9	9.4
2	6.2	86.8	10.4	2.8	82.3	5.2	12.5	12.2
3	7.3	89.2	7.9	2.8	81.5	4.7	13.8	10.9
4	8.3	91.2	6.6	2.2	80.4	5.0	14.6	12.8
5	9.1	90.3	6.2	3.5	79.4	5.2	15.4	12.4
6	9.7	92.7	4.5	2.8	77.9	6.0	16.1	14.1
7	10.7	91.6	4.7	3.7	78.4	5.1	16.5	14.0
8	12.0	92.5	3.3	4.3	76.7	6.1	17.1	17.5
9	14.0	91.3	3.0	5.7	76.2	5.6	18.2	14.2
10	18.4	88.3	2.3	9.4	71.7	8.6	19.7	21.8
All	100.0	90.1	5.2	4.7	77.7	6.0	16.4	15.1
Decile[b]	Share of Total Y_t	Composition of Y_t: E_t/Y_t	G_t/Y_t	M_t/Y_t	Uses of Y_t: C_t/Y_t	S_t/Y_t	T_t/Y_t	
B. Annual								
1	1.0	44.2	51.5	4.3	157.2	−62.4	5.2	
2	3.1	50.1	44.2	5.7	117.0	−21.1	4.1	
3	4.8	77.2	15.8	7.0	91.4	−0.3	8.9	
4	6.4	85.6	7.7	6.7	84.2	−4.0	11.9	
5	7.7	90.0	5.1	5.0	80.9	5.1	14.0	
6	9.0	89.8	4.3	5.9	77.6	7.1	15.4	
7	10.4	91.3	3.4	5.3	77.9	6.4	15.7	
8	12.1	90.2	2.6	7.2	76.3	7.2	16.5	
9	14.9	86.6	2.1	11.3	73.8	9.3	16.9	
10	30.6	57.5	1.2	41.4	49.2	34.4	16.4	
All	100.0	77.1	5.3	17.7	71.6	13.5	15.0	

Sources: E_t, G_t, T_t, E, G, and $\tilde{T}$ are based on data from the 1971 Statistics Canada Survey of Consumer Finances (*SCF*), as described in the text. All other variables are from the authors' simulations.

Notes: Part A: Gini coefficient of $L=.218$; Part B: Gini coefficient of $Y_t=.410$. Part A: L, E, G, I, C, B, $\tilde{T}$, and M are all present discounted lifetime values; L = lifetime resources, E = earnings, G = transfers, I = inheritances, C = consumption, B = bequest, $\tilde{T}$ = direct taxes on L, and M = investment income.

Part B: Y_t, E_t, G_t, M_t, C_t, S_t, and T_t are all annual flows; Y_t = income, E_t = earnings, G_t = transfers, M_t = investment income, C_t = consumption, S_t = saving, and T_t = direct taxes on Y_t.

[a]Shown in percent.

[b]Deciles are sorted by annual income as reported in the 1971 Statistics Canada *SCF*.

2) Over the lifetime transfers are less heavily concentrated in the bottom two deciles of the population than in the annual data. The decline in the relative importance of transfers as income rises is also less marked. This tends to reduce the quantitative importance of Browning and Johnson's argument regarding the impact of indexed transfers in a lifetime calculation.

3) Variation in consumption to income ratios is smaller in lifetime than in annual data. This reduces the regressivity of taxes assumed to be borne out of consumption, such as sales and excise.

4) There is less variation in the relative importance of earnings as we move up the income scale in lifetime data as against annual data. This is partly associated with the more even distribution of government transfers in the lifetime data. A consequence is that the progressivity of social security taxes over the low deciles in annual data is replaced by rough proportionality in the lifetime calculations. (In addition, regressivity over the upper deciles is reduced.) Also, the flatter earnings series again means that the Browning and Johnson procedure of allocating sales and excise taxes to factor incomes

has less tendency to produce progressivity for these taxes.

5) Overall, lifetime earnings are larger in relation to lifetime consumption than are annual earnings compared with annual consumption. This reflects the fact that, on average, earnings occur earlier in the lifetime and are less heavily discounted. We see in the next section that this factor tends to make overall lifetime incidence slightly more progressive than annual, since the relative importance of taxes falling on consumption is smaller and that of taxes on earnings is increased.

C. *Central Case Lifetime Incidence Assumptions*

The lifetime distributive series can be used in alternative ways to allocate the various taxes across deciles. Our central case uses standard competitive assumptions: the income tax is assumed to be borne by income recipients; corporate and property taxes by recipients of investment income; social security taxes by earnings recipients; and sales and excise taxes in proportion to consumption.

While for the most part it is clear from the central case incidence assumptions which distributive series should be used for which taxes, this is not true for taxes borne by capital in the lifetime context. In lifetime incidence, taxes on capital income have both sources and uses side effects rather than just a sources side impact as assumed in annual calculations.[19] By reducing the rate of interest they increase the relative lifetime incomes of those whose receipts occur later in life (via the discounting effect)—a sources side effect. In addition, they reduce the relative real incomes of persons who consume more later in the lifetime by increasing the price of future consumption—a uses side effect. While the sources and uses side effects of taxes borne by capital could be accounted for separately in the lifetime incidence calculations, it is simpler to net them out. This can be seen in the case of a family which consumes its total income in each year and is unaffected by taxes borne by capital, since it never saves and therefore feels no effect of a reduced interest rate. In contrast, a family which saves suffers from the reduced reward for abstinence.

Thus, taxes borne by capital reduce the welfare of families which defer consumption. In any year, in the absence of taxes borne by capital, a family which accumulated savings could consume more than with those same taxes present. The amount of additional consumption would equal the reduction in current investment income due to the taxes borne by capital. The appropriate distributive series for taxes borne by capital in lifetime incidence calculations is therefore the discounted value of all investment income received over the lifetime, M.

D. *Computing Tax Burdens*

As mentioned earlier, personal income tax payments are taken directly from the primary data source—the 1971 *SCF*. Social security tax payments are imputed using the applicable rates and ceilings under both the Canada Pension Plan (*CPP*) and Unemployment Insurance (*UI*) programs to determine employer and employee contributions on the basis of earnings reported in the 1971 *SCF* by simulation households. Other taxes must be imputed by establishing an overall tax rate and applying this to the appropriate distributive series.

For corporate and property taxes, rates are determined using 1970 tax collections and an estimate of total 1970 household investment income. We estimate the sales and excise tax rate by comparing actual 1970 tax collections with observed aggregate consumption. Lifetime capital and sales and excise tax burdens are calculated by applying these tax rates to discounted lifetime investment income M, and consumption C, respectively, for each simulation household.

[19]The nature of the uses side effect over the lifetime was pointed out by Browning and Johnson: "...corporation income taxes and property taxes reduce the net interest rate that savers receive. Consequently, these taxes harm savers and benefit consumers on the uses side..." (p. 27).

E. *Sensitivity Analyses*

In addition to lifetime incidence calculations under standard competitive assumptions, we present further results using a number of alternative assumptions to explore the robustness of our central case. We not only alter incidence assumptions, but also vary earnings mobility and other parameters in a sensitivity testing procedure. The calculations are kept on a comparable equal yield basis by using the aggregate tax collections from the central case,[20] and reallocating these by household according to the alternative assumptions used.

Our first variant on incidence assumptions is a "noncompetitive" version of the central case, comparable to that used in many annual studies. It is assumed that 50 percent of corporate and property taxes is shifted to consumers rather than to recipients of investment income. A second variant we refer to as the "Browning and Johnson Case." As discussed earlier, the argument of Browning and Johnson is that uses side effects of sales and excise taxes can be ignored and that transfers, being indexed, remain unaffected. The tax must therefore be borne by other sources of income. In this case, under competitive assumptions, sales and excise taxes are allocated using the distributions of lifetime earnings and inheritances instead of lifetime consumption.

Finally, we examine the sensitivity of our results to variations in key features of the life cycle model. We pay particular attention to the dependence of our results on the parameter which determines household mobility across the age specific distributions of earnings. Incidence results based on zero mobility (households occupy the same rank order in each age specific distribution) and free mobility (households' positions in each age distribution are random) are contrasted with the central case. We also rerun the life cycle simulation which generates our distributive series and recompute tax incidence using alternative long-run balanced growth rate assumptions of 1 and 4 percent, and interest rates of 4 and 8 percent. Results of these runs appear in the Appendix.

[20]The tax yield needs to be adjusted only in the cases where the interest rate or the rate of growth is altered.

IV. Lifetime Tax Incidence Results

Our central case lifetime tax incidence results are displayed in Table 2, Part A, with comparable annual tax incidence calculations reported in Part B. Several interesting points emerge from a comparison of these results.

First, the overall tax system appears mildly progressive in both the annual and lifetime calculations. While in the annual results there is regressivity from the bottom to the second decile, from the second to the ninth deciles rates increase slowly, and in the top decile a significant increase occurs. In the lifetime data there is progressivity throughout—with moderate increases in overall rates from the first to the fourth deciles, a slow rise from the fourth to the ninth deciles, and once more a larger increase for the top decile. While the regressivity at the bottom end in the annual run is an important feature, we conclude that both profiles show mild progression overall across the income ranges.

An alternative method of summarizing differences in tax profiles is to examine the proportional impact on a summary measure of income inequality. Following the suggestion of Musgrave and Tun Thin (1948), this is often done using the Gini coefficient.[21] The pre-tax Gini of .218 in the lifetime run is reduced to .184 by the tax system; in contrast, the annual Gini falls only 10 percent —from .410 to .370. On this basis it might be claimed that the lifetime tax structure is significantly more progressive than the annual.

The reason that our lifetime taxes give a larger proportional reduction in the Gini than in annual data, while the range of tax rates across the deciles is similar in both cases, lies in the smaller inequality in the pre-tax lifetime income distribution. Figure 1 provides

[21]Alternative approaches employing inequality indexes, intended to provide purely descriptive measures of overall progressivity were presented by Nanak Kakwani (1976) and Daniel Suits (1977). For a rigorous defense of the original Musgrave-Thin measure as a normative index and comments on Kakwani and Suits, see Pak-Wai Liu (forthcoming).

TABLE 2—AVERAGE TAX RATES OF CANADIAN HOUSEHOLDS BY DECILE, CENTRAL CASE ASSUMPTIONS

	1	2	3	4	5	6	7	8	9	10	All Deciles
A. Lifetime Incidence (Deciles Ranked by Lifetime Resources (L))											
Corporate Income Tax	2.2	2.9	2.6	3.0	2.9	3.3	3.3	4.1	3.4	5.1	3.6
Property Tax	2.4	3.1	2.8	3.3	3.2	3.6	3.6	4.5	3.7	5.6	3.9
Sales and Excises	15.0	14.3	14.1	13.9	13.8	13.5	13.6	13.3	13.2	12.4	13.5
Social Security	3.9	4.0	3.9	4.0	3.8	3.8	3.6	3.6	3.4	2.8	3.6
Personal Income Tax	7.3	11.3	12.5	13.5	14.5	15.1	15.7	16.7	17.7	20.5	15.8
All Taxes	30.9	35.5	35.9	37.7	38.1	39.3	39.8	42.2	41.3	46.5	40.2
B. Annual Incidence (Deciles Ranked by Annual Income (Y_t))											
Corporate Income Tax	1.0	1.3	1.6	1.6	1.2	1.4	1.2	1.7	2.7	9.8	4.2
Property Tax	1.1	1.5	1.8	1.7	1.3	1.5	1.3	1.9	2.9	10.6	4.5
Sales and Excises	27.2	20.3	15.8	14.6	14.0	13.4	13.5	13.2	12.8	8.5	12.4
Social Security	1.7	2.5	4.1	4.3	4.2	3.9	3.8	3.4	3.0	1.4	2.9
Personal Income Tax	4.3	2.8	6.8	9.7	11.9	13.4	13.8	14.8	15.4	15.7	13.5
All Taxes	35.4	28.4	30.1	31.9	32.6	33.6	33.7	35.0	36.8	46.0	37.5

Source: Computations performed by the authors. See text.

an illustration of how this works. Suppose that both annual and lifetime income are uniformly distributed, and that lifetime and annual tax profiles are both linear and have the same range. It is clear that the relative range of tax rates in the lifetime case is higher. Hence, for an index like the Gini which is concerned with relative differences in income, taxes appear more equalizing over the lifetime.[22]

A second interesting point evident from Table 2 is that there is considerably less variation in incidence patterns across taxes in the lifetime calculation. Lifetime income taxes while progressive are less so than in the annual case; sales and excise are less regressive; social security is initially less progressive and subsequently less regressive by income range, and corporate and property taxes are less progressive (especially at the top end). This reflects a greater flatness of the underlying lifetime distributive series and means that changes in incidence assumptions have less impact on lifetime, than on annual incidence results.

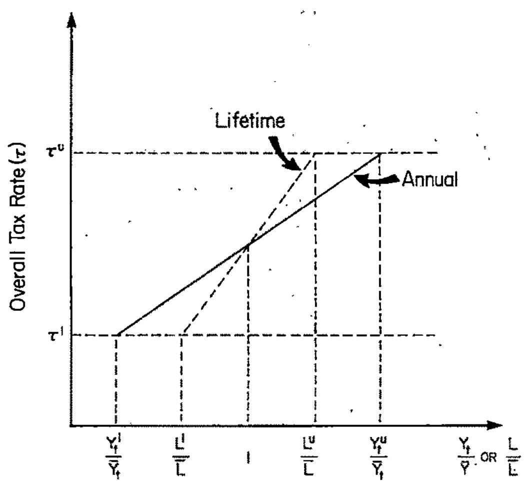


FIGURE 1. HYPOTHETICAL LIFETIME AND ANNUAL TAX PROFILES

Note: Y_t and L are annual and lifetime income, respectively. The bars indicate mean value; u and l superscripts denote upper and lower bounds, respectively.

Table 3 confirms this by showing how use of noncompetitive assumptions makes surprisingly little difference to overall tax rates. In this case, corporate and property taxes are treated as partially forward shifted and are allocated one-half to consumption and one-half to capital income in general. In this noncompetitive variant, overall progressivity

[22]We are indebted to an anonymous referee for suggesting the use of the Musgrave-Thin measure, and for providing the diagram in Figure 1. Since inequality in underlying income distributions frequently varies in incidence studies due to experiments with income definition, more general use of the Musgrave-Thin measure seems to us desirable.

TABLE 3—SENSITIVITY OF LIFETIME INCIDENCE RESULTS TO ALTERNATIVE INCIDENCE ASSUMPTIONS

	Deciles[a]										All
	1	2	3	4	5	6	7	8	9	10	Deciles
A. Central Case Results[a]											
Corporate Income Tax	2.2	2.9	2.6	3.0	2.9	3.3	3.3	4.1	3.4	5.1	3.6
Property Tax	2.4	3.1	2.8	3.3	3.2	3.6	3.6	4.5	3.7	5.6	3.9
Sales and Excises	15.0	14.3	14.1	13.9	13.8	13.5	13.6	13.3	13.2	12.4	13.5
Social Security	3.9	4.0	3.9	4.0	3.8	3.8	3.6	3.6	3.4	2.8	3.6
Personal Income Tax	7.3	11.3	12.5	13.5	14.5	15.1	15.7	16.7	17.7	20.5	15.8
All Taxes	30.9	35.5	35.9	37.7	38.1	39.3	39.8	42.2	41.3	46.5	40.2
B. Noncompetitive Assumptions											
Corporate Income Tax	3.1	3.3	3.2	3.4	3.3	3.4	3.4	3.8	3.4	4.2	3.6
Property Tax	3.4	3.6	3.4	3.6	3.9	3.7	3.7	4.2	3.7	4.6	3.9
Sales and Excises	15.0	14.3	14.1	13.9	13.8	13.5	13.6	13.3	13.2	12.4	13.5
Social Security	3.9	4.0	3.9	4.0	3.8	3.8	3.6	3.6	3.4	2.8	3.6
Personal Income Tax	7.3	11.3	12.5	13.5	14.5	15.1	15.7	16.7	17.7	20.5	15.8
All Taxes	32.8	36.4	37.1	38.4	39.2	39.5	40.1	41.6	41.4	44.6	40.2
C. Browning and Johnson Case											
Corporate Income Tax	2.2	2.9	2.6	3.0	2.9	3.3	3.3	4.1	3.4	5.1	3.6
Property Tax	2.4	3.1	2.8	3.3	3.2	3.6	3.6	4.5	3.7	5.6	3.9
Sales and Excises	12.0	12.7	13.1	13.3	13.3	13.6	13.5	13.7	13.8	13.9	13.5
Social Security	3.9	4.0	3.9	4.0	3.8	3.8	3.6	3.6	3.4	2.8	3.6
Personal Income Tax	7.3	11.3	12.5	13.5	14.5	15.1	15.7	16.7	17.7	20.5	15.8
All Taxes	27.9	34.0	34.9	37.0	37.6	39.4	39.8	42.6	41.8	48.0	40.2

Source: Computations performed by the authors. See text.
[a]Deciles ranked by lifetime resources.

of the tax system hardly changes at all. This is in sharp contrast to the results from annual studies where similar forward shifting assumptions produce marked reductions in progressivity.[23]

Finally Table 3 indicates that while Browning-Johnson assumptions make the tax profile more progressive, the change is small —markedly less than occurs in annual calculations. Denying the uses side effect of sales and excise taxes removed less regressivity in the lifetime than in annual calculations, because, over the lifetime, consumption is much flatter as a proportion of income across the income ranges. Recognizing the indexation of transfers increases progressivity less since transfers are less concentrated in the lower deciles in the lifetime data. The change in tax rates reported here is thus small when compared to the changes reported by Browning and Johnson. These results qualify the claim that changing the treatment of sales and excise taxes, as Browning and Johnson suggest transforms the overall incidence pattern from proportionality to sharp progressivity.

In Table 4 we report the impacts of changes in mobility assumptions on our incidence results. Not surprisingly, altering the degree of household mobility through the age-specific distributions of earnings changes measured lifetime inequality considerably. While in our central case, with mobility parameterized according to the best available empirical evidence (as explained earlier) we have a Gini coefficient for lifetime resources of .218, with fixed earnings ranks (zero mobility) the Gini rises to .326, and with completely random ("free") mobility it falls to .135. The altered inequality is reflected directly in the progressivity of personal income taxes. While in the central case the personal tax rate on a lifetime basis varies from 7 to 21 percent across deciles, with zero mobility

[23]See, for example, Pechman and Okner (Table 4.4, p. 51). Their variants 1a and 3b differ principally in that 3b shifts one-half of corporate and payroll taxes forward onto consumers (see p. 38).

TABLE 4—SENSITIVITY OF LIFETIME RESULTS TO ALTERNATIVE MOBILITY ASSUMPTIONS

	Percent of Total Lifetime Resources Accruing to Decile					
Deciles[a]	Zero Mobility Assumption		Central Case		Free Mobility Assumption	
A. Size Distribution of Lifetime Resources						
1	1.7		4.2		6.4	
2	4.1		6.2		7.7	
3	5.9		7.3		8.4	
4	7.3		8.3		8.9	
5	8.6		9.1		9.4	
6	9.9		9.7		9.9	
7	11.3		10.7		10.4	
8	12.8		12.0		11.1	
9	15.2		14.0		12.2	
10	23.2		18.4		15.5	
Gini Coefficient	.326		.218		.135	
B. Tax Rate Profiles	Income Tax	Total Taxes	Income Tax	Total Taxes	Income Tax	Total Taxes
1	2.1	28.9	7.3	30.9	12.7	36.7
2	5.2	30.8	11.3	35.5	13.1	37.7
3	8.4	33.7	12.5	35.9	14.1	38.0
4	11.1	35.8	13.5	37.7	14.1	37.2
5	13.2	38.7	14.5	38.1	15.7	40.5
6	14.0	38.3	15.1	39.3	15.6	40.5
7	14.9	39.7	15.7	39.8	15.2	39.1
8	16.6	42.7	16.7	42.2	15.8	40.4
9	17.4	41.4	17.7	41.3	17.4	43.1
10	22.7	47.4	20.5	46.5	18.8	46.1
All Deciles	15.8	40.7	15.8	40.2	15.8	40.6

Source: Computations performed by the authors. See text.
[a]Deciles ranked by lifetime resources.

the range widens from 2 to 23 percent, and with free mobility it narrows from 13 to 19 percent. Although the change in personal income tax progressivity has a limited effect on the overall incidence pattern,[24] it is clear that the assessment of lifetime tax incidence depends to a significant extent on the degree of mobility assumed.

Calculated overall progressivity of the tax system is relatively insensitive to the rate of interest and long-run balanced growth rate which are assumed, as shown in Table A1 in the Appendix. Choosing a higher growth rate slightly reduces progressivity, while the higher the interest rate the more progressivity we have. These effects operate via changes in the time paths of earnings relative to consumption. A higher growth rate skews earnings toward the future, reducing saving in early periods and therefore raising consumption and reducing investment income as a fraction of lifetime income. Sales and excise taxes (regressive) become more important and capital taxes (progressive) less. A higher interest rate works in the opposite direction, skewing consumption toward the future and increasing saving when households are young.

V. Conclusion

In this paper we report a set of lifetime tax incidence calculations which combine the data and output from a life cycle, microsimulation model for Canada due to Davies

[24]Offsetting changes occur in the incidence of the corporate and property, and sales and excise taxes. These are largely due to changes in saving patterns resulting from the altered age profiles of noninvestment income across the deciles.

(1979, 1982) with incidence assumptions comparable to those used in previous annual incidence studies. While incidence calculations have several limitations, previous annual calculations have received so much attention in policy debate that it is important to analyze to what extent they correspond to the more interesting case of lifetime tax incidence.

We offer two main findings. The first is that using standard "competitive" shifting assumptions, lifetime and annual incidence patterns both display mild progression of overall tax rates across income ranges (ignoring the bottom decile in the annual calculation). While the personal income tax becomes less progressive in a lifetime context, offsetting changes occur in the incidence of other taxes. Sales and excise taxes, for example, become less regressive since the fraction of income saved rises less sharply with income. The regressivity often found at the bottom end of the income scale in annual calculations is not present in the lifetime calculations. This is largely the result of the reduced regressivity of sales and excise taxes.

Our second finding is that lifetime incidence results are far more robust to alternative incidence assumptions than is true of annual results. The principal distributive series in the lifetime context—transfers, earnings, and consumption, as a fraction of lifetime income—are all closer to uniform across income ranges than corresponding annual series. As a result, changes in incidence assumptions such as the indexing of transfers (suggested by Browning and Johnson), or the forward shifting of capital taxes have much less impact on the overall progressivity of the tax system than in annual calculations. Our results thus indicate that we can perhaps be more confident than on the basis of annual incidence calculations alone that the incidence of the overall tax system is mildly progressive.

Several qualifications are in order, however. First, lifetime saving behavior by household is simulated rather than observed. While Davies (1982) found that the patterns of intergenerational saving generated are consistent with available evidence, our empirical knowledge of intergenerational transfers is not as well developed as that in other areas of consumer behavior. It would be valuable to compare our results with those obtained using alternative models of saving, or longitudinal data sets. Secondly, results are sensitive to the degree of earnings mobility assumed in our simulations. While mobility has been carefully parameterized with reference to studies using longitudinal data on earnings, more direct use of such data would be preferable.

Finally, the incidence assumptions behind our central (competitive) case are only correct under strong assumptions about factor supplies and technology. General equilibrium computations allowing differential incidence calculations under more interesting assumptions would clearly be more revealing. Adoption of the lifetime framework in detailed general equilibrium computations of tax incidence would seem to be a promising direction for future research to explore.

APPENDIX

Here we investigate the sensitivity of our central (competitive) case to changes in the rates of interest and growth rates which are assumed. Table A1 shows the overall lifetime incidence patterns obtained with alternative values of these parameters.

The table shows that overall progressivity falls slightly as higher growth rates are used, and rises somewhat with a higher rate of discount. Also, the average total tax rate falls with the rate of growth and rises with the rate of interest. The changes in progressivity are all fairly small, providing an indication that results of the paper would not be greatly altered with significantly different rates of interest and growth.

Raising the growth rate skews households' earnings more towards the future without altering the shape of their optimal consumption profiles significantly.[25] The result is lower saving at earlier ages, and a general

[25] There would be no change in shape (only in level) if households were not constrained to have nonnegative wealth. Skewing earnings more toward the future will result in this constraint being effective for more households.

TABLE A1—SENSITIVITY OF LIFETIME RESULTS TO THE CHOICE OF PARAMETERS

Deciles[a]	Total Tax Rate Profiles[b]		
A. Rate of Growth	1 Percent	2 Percent	4 Percent
1	31.6	30.9	29.8
2	36.5	35.5	33.6
3	36.9	35.9	34.8
4	39.8	37.7	35.7
5	38.6	38.1	36.8
6	42.0	39.3	36.8
7	42.1	39.8	37.8
8	44.8	42.2	38.6
9	43.7	41.3	38.9
10	49.5	46.5	43.4
All Deciles	42.3	40.2	37.9
B. Discount Rate	4 percent	6 Percent	8 Percent
1	29.0	30.9	33.5
2	33.2	35.5	38.4
3	33.5	35.9	39.0
4	34.7	37.7	41.1
5	35.6	38.1	41.0
6	36.2	39.3	44.7
7	36.0	39.8	44.2
8	37.4	42.2	47.2
9	38.3	41.3	45.4
10	42.1	46.5	52.1
All Deciles	36.9	40.2	44.5

Source: Computations performed by the authors. See text.
[a]Deciles ranked by lifetime resources.
[b]Part A: Assuming a rate of growth as shown; Part B: Assuming a discount rate as shown.

increase in discounted lifetime consumption as a fraction of lifetime income. There is also a general reduction in discounted lifetime investment incomes (due to lower saving when households are young).

The general rise in consumption as a fraction of lifetime income caused by increasing the growth rate, means that any given rate of sales and excise tax will translate into a higher tax as a proportion of lifetime income. Thus sales and excise taxes, which are fairly regressive (see Table 2) increase in importance. The general decline in lifetime investment income as a fraction of lifetime income similarly lowers corporate and property taxes as a proportion of lifetime income. In addition to the increased importance of the regressive sales and excise taxes, we thus have a somewhat reduced importance of progressive capital taxes (again see Table 2). It is therefore not surprising that overall progressivity declines somewhat. The slight decline in the average total tax rate is the result of capital taxes declining fairly sharply as a fraction of lifetime income.[26]

When we reduce the growth rate, the two mechanisms described work in reverse. We skew earnings more toward the present, and saving goes up considerably when households are young. Consumption falls and investment income rises as a fraction of lifetime income. Sales and excise taxes become less important, and capital taxes more. Overall progressivity therefore is somewhat increased.

Finally, as noted, Table A1 also shows that a higher rate of discount raises progres-

[26]In the central case (growth rate = 2 percent) capital taxes average 7.5 percent of lifetime income, and other taxes 32.5 percent. In the variant we are considering with a growth rate of 4 percent, capital taxes average 4.5 percent, and other taxes 33.4 percent of lifetime income.

sivity, and the average total tax rate slightly. The explanation runs along lines similar to that for changes in the growth rate. A higher rate of interest skews planned consumption more toward the future, without altering the time path of earnings. Saving in early periods therefore goes up, leading to a decline in consumption, and a rise in investment income as a proportion of lifetime income. Sales and excise taxes become less important, and capital taxes more. Overall progressivity therefore rises. The increase in average total tax rates is somewhat larger than when the growth rate is reduced, however. The reason is that the rise in investment income as a fraction of lifetime income is greater than when the growth rate is reduced, leading to a larger rise in capital taxes as a proportion of lifetime income.[27]

[27]When we raise the interest rate to 8 percent, on average discounted lifetime investment income rises to 23.6 percent of lifetime income, whereas reducing the growth rate to 1 percent only produces a rise to 20.1 percent. (Investment income averages 15.1 percent of lifetime income in the central case—see Table 1.) The result is that capital taxes rise to 11.7 percent on average, from 7.5 percent, when the interest rate is raised, as compared with a rise to 9.8 percent when the growth rate is reduced to 1 percent.

REFERENCES

Atkinson, Anthony B. and Stiglitz, Joseph E., *Lectures in Public Economics*, London: McGraw-Hill, 1980.

Blinder, Alan S., *Toward an Economic Theory of Income Distribution*, Cambridge: MIT Press, 1974.

Blomqvist, Nils S., "A Comparison of Distributions of Annual and Lifetime Income: Sweden Around 1970," *Review of Income and Wealth*, September 1981, *27*, 243–64.

Browning, Edgar K., "The Burden of Taxation," *Journal of Political Economy*, August 1978, *86*, 649–71.

______ **and Johnson, William R.,** *The Distribution of the Tax Burden*, Washington: American Enterprise Institute, 1979.

Davies, David G., "An Empirical Test of Sales-Tax Regressivity," *Journal of Political Economy*, February 1959, *67*, 72–78.

Davies, James B., (1979a) "Life-Cycle Savings, Inheritance, and the Distribution of Personal Income and Wealth in Canada," unpublished doctoral dissertation, London School of Economics, 1979.

______, (1979b) "On the Size Distribution of Wealth in Canada," *Review of Income and Wealth*, September 1979, *25*, 237–59.

______, "The Relative Impact of Inheritance and Other Factors on Economic Inequality," *Quarterly Journal of Economics*, August 1982, *97*, 471–98.

Gillespie, W. Irwin, "On the Redistribution of Income in Canada," *Canadian Tax Journal*, July-August 1976, *24*, 419–50.

______, *The Redistribution of Income in Canada*, Ottawa: Gage Publishing, Ltd., 1980.

Kakwani, Nanak C., "Measurement of Tax Progressivity: An International Comparison," *Economic Journal*, March 1976, *87*, 71–80.

Kotlikoff, Laurence J. and Summers, Lawrence H., "The Role of Intergenerational Transfers in Aggregate Capital Accumulation," *Journal of Political Economy*, August 1981, *89*, 706–32.

Lillard, Lee A., "Inequality: Earnings vs. Human Wealth," *American Economic Review*, March 1977, *67*, 42–53.

Liu, Pak-Wai, "Lorenz Domination and Global Tax Progressivity," *Canadian Journal of Economics*, forthcoming.

Mincer, Jacob, *Schooling, Experience and Earnings*, New York: Columbia University Press, 1974.

Musgrave, Richard A. and Musgrave, Peggy B., *Public Finance in Theory and Practice*, New York: McGraw-Hill, 1980.

______, **Case, Karl E. and Leonard, Herman B.,** "The Distribution of Fiscal Burdens and Benefits," *Public Finance Quarterly*, July 1974, *2*, 259–311.

______ **and Thin, Tun,** "Income Tax Progression, 1929–48," *Journal of Political Economy*, December 1948, *56*, 498–514.

Pechman, Joseph A. and Okner, Benjamin A., *Who Bears the Tax Burden?*, Washington: The Brookings Institution, 1974.

St-Hilaire, France and Whalley, John, "What Have We Learned from Tax Incidence Calculations?," mimeo., University of

Western Ontario, 1982.

______ and ______, "A Microconsistent Equilibrium Data Set for Canada for Use in Tax Policy Analysis," *Review of Income and Wealth*, June 1983, *29*, 175–204.

Suits, Daniel B., "Measurement of Tax Progressivity," *American Economic Review*, September 1977, *67*, 747–52.

Statistics Canada Survey of Consumer Finances (*SCF*), *1971 Income Distribution by Size in Canada*, Publication No. 13-207, Ottawa: Ministry of Supply and Services, 1973.

Two Agency-Cost Explanations of Dividends

By Frank H. Easterbrook*

The economic literature about dividends usually assumes that managers are perfect agents of investors, and it seeks to determine why these agents pay dividends. Other literature about the firm assumes that managers are imperfect agents and inquires how managers' interests may be aligned with shareholders' interests. These two lines of inquiry rarely meet.[1] Yet logically any dividend policy (or any other corporate policy) should be designed to minimize the sum of capital, agency, and taxation costs. The purpose of this paper is to ask whether dividends are a method of aligning managers' interests with those of investors. It offers agency-cost explanations of dividends.

I. The Dividend Problem

Businesses find dividends obvious. Boards declare them regularly and raise them from time to time or face disquiet from investors, or so they think. Many managers are sure that higher dividends mean higher prices for their shares. There is a substantial body of law that controls when boards may (sometimes must) declare dividends, in what amount, and using what procedures.[2] Firms enter into complicated contracts with creditors and preferred stockholders that govern the permissible rates of payouts.[3] Dividends are paid (and regulated) at considerable cost to the firms involved.

Economists find dividends mysterious. The celebrated articles by Merton Miller and Franco Modigliani declared them irrelevant because investors could home brew their own dividends by selling from or borrowing against their portfolios. Meanwhile the firms that issued the dividends would also incur costs to float new securities to maintain their optimal investment policies.[4] Dividends are, moreover, taxable to many investors, while firms can reduce taxes by holding and reinvesting their profits. Although dividends might make sense in connection with a change in investment policy—when, for example, the firms are disinvesting because they are liquidating or, for other reasons, shareholders can make better use of the money than managers—they are all cost and no benefit in the remaining cases of invariant investment policies.[5]

Dividends are hard enough to explain when they occur in isolation; a combination of dividends and simultaneous raising of new capital is downright inexplicable.[6] Yet the simultaneous or near-simultaneous payment of dividends and raising of new capital are common in business. Sometimes firms issue new stock at or around the time they pay

*University of Chicago Law School, 1111 East 60th Street, Chicago, IL 60637. I thank Douglas Baird, Walter Blum, Dennis Carlton, Harry DeAngelo, Eugene Fama, Daniel Fischel, Michael Jensen, William Landes, Saul Levmore, Merton Miller, Myron Scholes, and Alan Schwartz for helpful comments on earlier drafts.

[1] One meeting place is Michael Rozeff (1982), who suggests that dividends and agency costs are related and offers a test showing that dividends depend in part on the fraction of equity held by insiders. He does not provide any mechanism, however, by which dividends and the consequent raising of capital control agency costs. I discuss some mechanisms of this sort below.

[2] See Bayless Manning (1981), for a description of the legal rules. Compare Victor Brudney (1980) calling for more legal regulation with Daniel Fischel (1981) offering economic support for current legal rules.

[3] See Clifford Smith and Jerold Warner (1979); Kose John and Avner Kalay (1982).

[4] See Miller and Modigliani (1961); also Modigliani and Miller (1958); Joseph Stiglitz (1974).

[5] See Modigliani (1982), for an argument to this effect that modifies the M-M irrelevance model by considering taxes and uncertainty. Compare Miller and Myron Scholes (1982), applying their earlier analysis (1978) to extend the irrelevance hypothesis to a world with taxes. Miller and Scholes argue that taxes need not, and do not appear to, determine dividend policy; all the same, their analysis does not show why there are dividends in a world of costly flotation.

[6] For example, Miller and Kevin Rock: "It would be uneconomic as well as pointless" for firms to pay dividends and raise capital simultaneously (1982, p. 13).

dividends. More frequently they issue new debt, often in the form of bank loans that are almost invisible to finance economists. Why does this occur?

The problem with the irrelevance proposition is that dividends are costly yet ubiquitous. Something causes them. Even if most investors are irrational most of the time, dividends would go away if their costs exceeded their benefits to investors. Firms that reduced payouts would prosper relative to others; investors who figured out the truth also would prosper relative to others; and before long—certainly before now in light of the large costs of floating new capital issues and the large differences between income and capital gains tax rates—dividends would be infrequent occurrences characterizing failing or disinvesting firms.

The existence of dividends despite their costs has inspired a search for explanations. Some of the efforts have been obvious failures. Take the argument that investments are risky and that dividends hedge against the possibility that the firms will go bankrupt before distributing the saved-up assets to the shareholders. The argument goes: investors value a steady stream of dividends over the uncertain prospect of a large return when the firms liquidate or are sold as going concerns and the investors are cashed out, and firms pay dividends to cater to that preference. The problem here is that dividends are matched by reinvestments: so long as dividends do not affect firms' investment policies, they do not represent any withdrawal of capital from risky ventures. New investors bear the risk that the dividend-receiving investors avoid, and these new investors must be compensated. The new investors may well turn out to be the old ones; shareholders do not usually use the dividends for consumption or to purchase Treasury bills. If they reinvest the proceeds in the same or a different firm, they commit their cash (less taxes paid) to the same risks as if there had been no dividends. In sum, there is no bird-in-the-hand effect unless the firm also changes its investment policy.[7]

Other arguments are only slightly more plausible. Consider the argument that dividends "signal" the well-being of the firm to investors and so promote confidence (and, one supposes, higher stock prices and a flow of investment capital).[8] The problem here is that it is unclear just what dividends signal, how they do so, or why dividends are better signals than apparently cheaper methods. Firms could and do issue disclosures of their prospects and profits. True, investors may be disinclined to believe the self-serving statements of managers about the firms' endeavors, but managers' usual response to this is to hire outsiders who examine the firms' books and other materials and opine on whether the managers are telling the truth. These outsiders work for many firms and acquire reputational capital so large that they become unbribable. No firm could offer them enough for a false (or slipshod) verification to make up for their losses on business with other firms. Auditors serve this function yearly or more often;[9] even judges may serve this function in suits charging the managers with making false statements or omitting material facts.

Dividends would be desirable only if they added to the efficacy of these methods of disclosure. The beauty of a "signal" is that it is self-verifying. People believe the signal because sending the message is rational for the signaller only if the message is or is believed to be accurate. Thus one could say that a Ph.D. from the University of Chicago is a good signal of intelligence and diligence (two notoriously hard-to-verify qualities) be-

[7]Compare M. J. Gordon (1959), with Michael Brennan (1971) and Sudipto Bhattacharya (1979). It is sometimes said that the bird-in-the-hand argument fails because one may get cash for consumption (or to put in the bank) by selling stock as well as by waiting for dividends. This is not a good refutation, because if the lack of dividends puts invested dollars at unacceptable risk, shares will fetch less in the market on a no-dividend policy than otherwise, and investors will be poorer than they would be if dividends were plentiful.

[8]See, for example, Bhattacharya; Nils Hakansson (1982); Steven Ross (1977). Compare Miller and Rock: in this model, dividends may permit inference of sources and uses of funds; this achieves many of the effects of a signalling model but by direct revelation or inference.

[9]See Linda DeAngelo (1981): using the auditor's "reputational bond" to show that larger auditors provide better quality.

cause persons of inferior intellect could not obtain one. But dividends do not directly reveal the prospects of the firms, so the message they convey may be ambiguous.[10] Unless the cost of issuing dividends is uniformly lower for prosperous firms, no signal is possible.

Prosperous firms may withhold dividends because internal financing is cheaper than issuing dividends and floating new securities. Worse, dividends do not distinguish well-managed, prospering firms from others. They are not irrational for poorly managed or failing firms. Quite the contrary, such firms should disinvest or liquidate, and their managers may choose dividends as a method of accomplishing this. Someone who observes an increase in the dividend has no very good way to tell whether this signals good times or bad. (This is consistent with both the finding that dividends are poor predictors of future net earnings and the finding that stock prices are poor predictors of future dividends.)[11] Doubtless only a prospering firm can continue to pay dividends year in and year out, but a firm with a long record of prosperity also would not need the verification available from the dividend signal. The persistent reports of auditors and securities dealers, its securities' prices, and the apparent marketing success of the firm would do as well in verifying the managers' tales.

The explanations based on clientele effects also are unsatisfactory. It is easy enough to see that if some investors are in different tax positions from others (for example, some hold tax-sheltered funds while others are taxed at ordinary rates), the different groups will have different preferences for dividends. The taxed group would prefer to take profits as capital gains; the untaxed group would be indifferent. Some equilibrium would develop in which firms adopted different payout policies to cater to the different clienteles.[12] It is much harder, though, to use clientele effects to demonstrate why the current dividend structure exists. Why do most firms pay significant dividends, given the costs of paying them (and raising new capital), and given that all investors either prefer capital gains or are indifferent between dividends and capital gains?[13]

II. Two Explanations

The dividend puzzle has been stated as: "what is the effect of a change in cash dividends paid, *given the firm's capital-budgeting and borrowing decisions*?"[14] This statement of the problem makes it insoluble, because the irrelevance hypothesis and the growing body of evidence say that dividends do not matter so long as the firm's financing and investment policy is unaffected. The existence of dividends in the face of this, and despite the costs of paying them out and raising new money, suggest that it is appropriate to ask a different question: "what is the effect of a *consistent* policy of paying dividends?" This question leads to what could be called a naive explanation of dividends. Dividends exist *because* they influence the firms' financing policies, *because* they dissipate cash and induce firms to float new securities.

Let us suppose that managers are not perfect agents of the other participants in the corporate venture, but that they pursue their own interests when they can. Because the managers are not the residual claimants to the firm's income stream, there may be a substantial divergence between their interests and those of the other participants. Man-

[10] Only "may be" instead of "is" because the message may be self-justifying, as in the Miller and Rock model.

[11] Stephen Penman (1983) finds that knowledge of dividends adds little or nothing to earnings forecasts as predictors of future earnings. See also, for example, Robert Shiller (1981).

[12] See Fischer Black and Scholes (1974) and Alan Auerbach (1982). But see Miller and Scholes (1982).

[13] Martin Feldstein and Jerry Green (1983) do not overcome this problem with clientele models. They use a two-firm, two-holder model in which portfolio diversification, without steady trading, depends on differential dividend policies. The two-firm assumption drives the model. With large numbers of firms an investor can get plenty of diversification without differential dividend policies.

[14] Richard Brealey and Stewart Myers (1981, p. 324, italics in original).

agers, investors, and other participants will find it advantageous to set up devices, including monitoring, bonding, and *ex post* readjustments that give managers the incentive to act as better agents. The costs of monitoring, bonding, and the residual losses from slippage are agency costs borne by investors.[15]

One form of agency cost is the cost of monitoring of managers. This is costly for shareholders, and the problem of collective action ensures that shareholders undertake too little of it.[16] Although a monitor-shareholder would incur the full costs of monitoring, he would reap gains only in proportion to his holdings. Because shares are widely held, no one shareholder can capture even a little of the gain. Shareholders would be wealthier if there were some person, comparable to the bondholders' indenture trustee, who monitored managers on shareholders' behalf.

A second source of agency costs is risk aversion on the part of managers.[17] The investors, with diversified portfolios of stocks, will be concerned only about any nondiversifiable risk with respect to any one firm's ventures. Managers, though, have a substantial part of their personal wealth tied up in their firms. If the firms do poorly or, worse, go bankrupt, the managers will lose their jobs and any wealth tied up in their firms' stock. Managers therefore will be concerned about total risk, and their personal risk aversion will magnify this concern.

The risk-averse managers may choose projects that are safe but have a lower expected return than riskier ventures. Shareholders have the opposite preference. Riskier ventures enrich shareholders at the expense of creditors (because shareholders do not pay any of the gains to bondholders, yet bondholders bear part of the risk of failure), and shareholders would want managers to behave as risk preferrers.[18] Of course, creditors recognize this and try to control it in advance through bond indentures and other instruments; they also adjust the rate of interest they demand. Debtholders assume that given the limits set by their contracts, shareholders prefer to take the maximum advantage. But the question is not whether the riskiness of projects can be controlled through indentures or other legal devices. It is, rather, whether costs of control (including the costs of control and residual agency costs) can be reduced by a method that includes dividends.

Managers can change the risk of the firm not only by altering its mix of projects, but also by altering its debt-equity ratio. The lower the ratio of debt to equity, the lower the chance of bankruptcy of the firm. Once again, debtholders consider this in deciding what rate of interest to demand. Once again, given the existence of debt, managers can control the amount of risk. One way they can do this is by picking a dividend policy. If managers first issue debt and then finance new projects out of retained earnings, the debt-equity ratio will fall. The lower it falls, the lower the managers' risk and the greater the boon bestowed on the debtholders, who receive their contracted-for interest but escape the contracted-for risk. Financing projects out of retained earnings—if unanticipated by bondholders—transfers wealth from shareholders to debtholders. Just as bondholders want to limit dividends, to prevent advantage-taking by shareholders once a rate of interest has been set, so shareholders want to increase dividends to the extent possible in order to avoid being taken advantage of by bondholders.[19]

Shareholders therefore would like to induce managers to take more risks, so that

[15]Michael Jensen and William Meckling (1976); Eugene Fama (1980); Bengt Holmstrom (1982); *JLE* Symposium (1983).

[16]See my articles with Fischel (1982, 1983a) for discussions of the extent to which legal rules address this problem.

[17]See Jensen and Meckling (pp. 349–50, 352–53), Steven Shavell (1979), and Alan Marcus (1982).

[18]This is one possible argument for permitting insider trading, if other compensation schemes are too costly. Compare Henry Manne (1966) and Dennis Carlton and Fischel (1983), with my article (1981, pp. 330–38).

[19]John and Kalay also stress this, See also George Handjinicolaou and Kalay (1982): stressing role of dividends in adjusting risk between bondholders and stockholders; concluding that dividends do not appear to cause unanticipated losses to bondholders.

they do not give away wealth to bondholders. The shareholders would prefer that managers go to the limit authorized by contract in imposing risks on the firm's creditors. Yet it is hard to give managers the right incentives to do this. There is little one can do to get rid of their risk aversion. They will remain undiversified, because of the nature of their human capital, no matter what; indeed, the lack of diversification in managers' holdings has other benefits.[20] Unless there is some form of *ex post* settling up with managers, which will be difficult (costly) to achieve, shareholders' payoffs will be lower, with consequences for the level of investment.

Both the monitoring problem and the risk-aversion problem are less serious if the firm is constantly in the market for new capital. When it issues new securities, the firm's affairs will be reviewed by an investment banker or some similar intermediary acting as a monitor for the collective interest of shareholders, and by the purchasers of the new instruments. The same occurs when the firm issues new debt, including bonds, commercial paper, and syndicated bank loans. Managers who need to raise money consistently are more likely to act in investors' interests than managers who are immune from this kind of scrutiny. Moreover, when it issues new securities, the firm can adjust the debt-equity ratio (and obtain a new rate of interest for its debt) so that neither shareholders nor bondholders are taking advantage of the other group. (It can, of course, make this adjustment in other ways, including making more frequent trips to financial markets for smaller sums of new cash, but because flotation costs decrease with the size of the offering, such alternatives may be more costly than combining infrequent flotation with dividends.)

The principal value of keeping firms constantly in the market for capital is that the contributors of capital are very good monitors of managers. The firm's existing investors can influence the managers' actions only by voting (which suffers from a collective choice problem) and by selling. Purchasers of stock will pay no more than the value of future profits under current management unless they are prepared to wage a takeover contest of some sort, which can be very costly. Managers of firms with fixed capital structures may well have substantial discretion to be slothful, consume perquisites, or otherwise behave in their own interests rather than the investors' interests.

New investors do not suffer under the collective choice disabilities of existing investors. They can examine managers' behavior before investing, and they will not buy new stock unless they are offered compensation (in the form of reduced prices) for any remediable agency costs of management. Managers who are in the capital market thus have incentives to reduce those agency costs in order to collect the highest possible price for their new instruments. New investors are better than old ones at chiseling down agency costs.

Of course, new investors need information, and that may be hard to come by. Neither auditors nor the managers themselves are perfectly reliable unless there is a foolproof legal remedy for fraud.[21] Other forms of information gathering, such as shareholders' inquiries and stock brokers' studies, suffer from the problem that none of the persons making inquiry can capture very much of the gain of this endeavor, and thus there will be too little information gathered. There would be savings if some information gatherers had larger proportionate stakes, and if the verification of information could be accomplished at lower cost. Underwriters of stock and large lenders may supply the lower-cost verification. These firms put their own money on the line, and any information inferred from this risk-taking behavior by third parties may be very valuable to other investors. This form of verification by acceptance of

[20] Douglas Diamond and Robert Verrecchia (1982). Fama and Jensen (1983a) offer a substantially different treatment, in which they do not employ the artifact of the firm as risk-averse "principal." Managers then bear risk in the form of undiversified portfolios in order to induce reductions in other agency costs.

[21] See Stanford Grossman (1981); myself and Fischel (1983b).

risk is one of the savings that arise when dividends keep firms in the capital market.[22]

The role of dividends in starting up the monitoring provided by the capital market is easy to see. An example of the role of dividends in making risk adjustments may help. Suppose a firm has an initial capitilization of 100, of which 50 is debt and 50 equity. It invests the 100 in a project. The firm prospers, and earnings raise its holdings to 200. The creditors now have substantially more security than they started with, and correspondingly the residual claimants are paying the creditors a rate of interest unwarranted by current circumstances. They can correct this situation by paying a dividend of 50 while issuing new debt worth 50. The firm's capital continues to be 200, but the debt-equity ratio has been restored, and the interest rate on the original debt is again appropriate to the creditors' risk.[23]

Expected, continuing dividends compel firms to raise new money in order to carry out their activities. They therefore precipitate the monitoring and debt-equity adjustments that benefit stockholders. Moreover, even when dividends are not accompanied by the raising of new capital, they at least increase the debt-equity ratio so that shareholders are not giving (as much) wealth away to bondholders. In other words, dividends set in motion mechanisms that reduce the agency costs of management and that prevent one group of investors from gaining, relative to another, by changes in the firm's fortunes after financial instruments have been issued.[24] The future is always anticipated imperfectly in these contracts, so there will always be some need for *ex post* adjustments and supervision, and dividends play a role in these adjustments.

This obviously is not altogether different from information or signalling explanations of dividends. One could recharacterize part (but not all) of this treatment as an assertion that investment bankers and other financial intermediaries send signals to investors by putting their reputations (and, in underwritten offerings, money) on the line and certifying that the new securities are backed by the represented earnings potential. The information interpretation of this agency-cost treatment at least offers a plausible explanation why dividends (rather than, say, earnings announcements) carry essential information.

There is a further problem because the explanations I have offered are not unique explanations *of dividends*. Nothing here suggests that repurchases of shares would not do as well as or better than dividends. The issuance of debt instruments in series, so that payments and refinancings are continuous, serves the same function as dividends. I have "explained" only mechanisms that keep firms in the capital market in ways that instigate consistent monitoring and consistent readjustment of the risk among investors.

The explanation I have offered also is open to the objection, along the lines of Fischer Black (1976), that shareholder-creditor conflicts may be resolved by negotiation after any change in the fortunes of the firm. The investors could agree to new payoffs or shares of control rather than to a dividend policy. This may well be true, but such *ex post* negotiation raises a bilateral monopoly problem, and the costs of the negotiation could be substantial even if (contrary to my assumption) there were no agency problems. Unless *ex post* negotiation is very costly, the

[22] One thus cannot treat it as paradoxical that in raising capital firms use investment bankers at a cost greater than the firm would incur in raising capital via rights offerings or other non-intermediated devices. Compare Robert Hansen and John Pinkerton (1982).

[23] Some cases contain an implicit recognition of this function of dividends. For example, *Randall v. Bailey*, 23 N.Y.S.2d 173 (N.Y. Sup. Ct. 1940), aff'd mem., 262 App. Div. 844, 29 N.Y.S.2d 512 (1st Dep't 1941), aff'd, 288 N.Y. 280, 43 N.E.2d 43 (1942) (permitting dividend out of unrealized appreciation, financed by new debt).

[24] This explanation of dividends is closely related to the one Grossman and Oliver Hart (1982) offer for debt. They say that debt is desirable to equity holders precisely because it creates bankruptcy costs, thus inducing managers to take extra care. I say, in parallel fashion, that dividends are beneficial to equity holders because they force managers constantly to obtain new capital in competitive markets. Fama and Jensen (1983b, pp. 13–15) also treat debt as a mechanism for regulating agency costs, although their argument is not the same as that of Grossman and Hart. See also Saul Levmore (1982, pp. 70–71). There is a family resemblance among all of these arguments.

existing pattern of complex bond indentures that provide for most contingencies makes little sense. I therefore think we must assume that in some decently large number of cases, accommodation through dividends and financing decisions set by the residual claimants is cheaper than accommodation through *ex post* negotiation. This is, however, an empirical matter, which raises the question whether the agency-cost explanations are testable.

III. Possible Tests

There have been a flurry of tests on the consequences of dividends. Some show that dividend changes are not related to the price of shares; others claim that increases in dividends are associated with decreases in the prices of shares.[25]

These are hard to evaluate, for it is hard to obtain a measure of unanticipated changes in the level of dividends, and only unanticipated changes could change the prices of shares. The "level of dividends" is itself difficult to calculate for purposes of these studies. Earlier treatments of dividends seek to explain *net* dividends (payouts in excess of new flotation), and it is almost impossible to obtain data on net dividends. Moreover, because an increase in dividends could be caused either by an increase in the firm's profits (implying higher stock prices) or by the commencement of disinvestment as the firm has fewer profitable opportunities (implying lower stock prices), studies that aggregate dividend increases across the classes could show small effects even when studies separating the two reasons for increase would show large ones. The studies have other problems as well.

It is not my purpose to offer a critique of the available work. Any study of the implications of the hypotheses I offer here would be beset by many of the same difficulties. Because of the agency-cost explanations of dividends focus on constant payout policies rather than changes in dividends, new tests will raise difficult questions of anticipation.

It should be possible to reexamine the data using as a new independent variable whether the firm had been in the capital market raising new money (whether debt or equity) at much the same time as it was paying dividends. The presence of new fund raising would indicate that dividends did not represent disinvestment. It also would isolate the set of firms whose managers were not able to rely wholly on internally generated funds and for which, therefore, dividends might reduce agency costs. The hypotheses I offer here suggest that the securities of firms simultaneously paying dividends and raising new money will appreciate relative to other securities. A test, however, will encounter substantial difficulty in identifying the time at which new capital is raised. Some syndicated loans are not announced to the public, and what is one to make of a firm's drawing against a line of credit arranged at an earlier time? A test also would face the problem of determining which payout policies were anticipated and which were not.

Finding a significant number of firms simultaneously (another problem of interpretation!) paying dividends and raising new money should offer substantial support for the agency-cost explanations, because other approaches to dividends imply that firms raise capital or pay dividends but do not do both. The agency-cost explanations also offer a plausible reason, other than clientele effects, why changes in dividend or financing policy may be associated with price reductions. It would be interesting to find out whether simultaneous dividend and financing changes produce the same negative residuals sometimes found when one changes but the other does not. One also could attempt to distinguish firms with high dividend/financing-to-asset ratios from firms with low dividend/financing-to-asset ratios. The hypotheses offered here suggest that there is some optimal ratio for each firm or set of firms.[26]

The difficulty of designing an empirical test is formidable, which suggests resorting

[25] For example, Robert Litzenberger and Krishna Ramaswamy (1982); Miller and Scholes (1982); Miller and Rock; all discussing earlier studies.

[26] Existing findings that new financings are associated with price reductions do not account for the possibility that rights offerings and below-market underwritten offerings will reduce the price of stock without diluting the current investors' interests (and thus without making them worse off).

to some less formal inquiries. The agency-cost explanations of dividends imply that dividends are worthless in themselves. Thus if firms are driven to the capital market by other conditions, we would expect to see less paid out in dividends. This is consistent with the observation that no-dividend (or low-dividend) stocks are usually "growth" firms, which are regularly in the capital market, and with the impression that such firms start paying dividends only when the rate of their growth (and thus the frequency of their trips to the capital market) has been reduced. The need to find some agency-cost control device increases as a firm becomes older and the original devices are less well adapted to the current form of business, and the initiation of dividends may supply such a device from the capital market.[27]

The agency-cost explanations have implications for the stability of dividends over time as well. Because the first function of dividends is to keep firms in the capital markets, we would not expect to see a very strong correlation between short-term profits and dividends. This implication cries out for testing, but it is certainly consistent with the fact that most firms have consistent policies (for example, 20 cents per share per quarter) that are not changed very often. A consistent policy uncouples dividends from profits while maintaining a link to the capital market. One indirect way to examine whether consistent dividends are valued for their effect on agency costs is to examine whether prices appreciate more on an increase in the "regular" dividend than on an increase of the same present value in "extraordinary" dividends. Shareholders concerned only about payouts in hand value the two equally; if dividends contain agency costs, regular payouts are more valuable. Evidence indicates that the regular dividend is associated with greater increases in price (see James Brickley, 1983).

Profits would have some effect on the risk-adjustment function, but past profits (which inure to debtholders' benefit unless dividends are increased) would be more important explanatory variables than current or anticipated profits. Anticipated profits can be handled by an adjustment of the terms on which money is raised; unanticipated past profits must be paid out to avoid windfalls to debtholders. The agency-cost treatment predicts that increases in dividends lag increases in profit and are uncorrelated with future profits. The lag may be substantial, because small increases (small changes of all sorts) in profits will be anticipated by debtholders, and there will be no need to make adjustments for these changes. Only the unanticipated (relatively large) changes call for adjustments in dividend policy.

Finally, because all forms of controlling agency costs are themselves costly, we would expect to see substitution among agency-cost control devices. One method of dealing with agency costs is for the managers to hold substantial residual claims in the firm. As such managers' claims increase, other things equal, dividends would be less valuable to investors and would decrease. Michael Rozeff suggests that this occurs. The same sort of substitution should accompany use of other devices.

The agency-cost explanations do not, however, yield unambiguous predictions about how bond prices will react to dividends. On the one hand, dividends favor investors and expose bondholders to more risk, thereby depressing bond prices. Of course, bondholders anticipate the use of dividends and the ensuing adjustment of risk, so it is not clear whether any price effects will be large. On the other hand, dividends keep managers' noses to the grindstone, conferring benefits on all investors. The net effect may be a wash, or it may not.[28]

IV. Conclusion

The economics literature has yet to integrate the study of corporate finance and the theory of the firm. This paper is a small step toward understanding whether, and how, dividends may be useful in reducing the agency costs of management. I suggest that dividends may keep firms in the capital

[27]Paul Asquith and David Mullins (1983) find that there is a significant appreciation in the price of stock when a firm initiates dividend payments. This offers some support for the thesis in the text. (Asquith and Mullins treat the increase as an information effect; they do not consider other explanations for their findings.)

[28]Handjinicolaou and Kalay show ambiguous effects.

market, where monitoring of managers is available at lower cost, and may be useful in adjusting the level of risk taken by managers and the different classes of investors. Such an explanation offers a hope of understanding why firms simultaneously pay out dividends and raise new funds in the capital market. It does not, however, explain dividends (as opposed to the set of all devices that have the effect of keeping firms in the capital market), and it will be difficult to test.

REFERENCES

Asquith, Paul and Mullins, David W., "The Impact of Initiating Dividend Payments on Shareholders' Wealth," *Journal of Business*, January 1983, *56*, 77–96.

Auerbach, Alan J., "Stockholder Tax Rates and Firm Attributes," Working Paper No. 817, National Bureau of Economic Research, 1982.

Bhattacharya, Sudipto, "Imperfect Information, Dividend Policy, and 'The Bird in the Hand Fallacy'," *Bell Journal of Economics*, Spring 1979, *10*, 259–70.

Black, Fischer, "The Dividend Puzzle," *Journal of Portfolio Management*, Winter 1976, *2*, 5–8.

______ **and Scholes, Myron S.,** "The Effects of Dividend Yield and Dividend Policy on Common Stock Prices and Returns," *Journal of Financial Economics*, May 1974, *1*, 1–22.

Brealey, Richard and Myers, Stewart, *Principles of Corporate Finance*, New York: McGraw-Hill, 1981.

Brennan, Michael J., "A Note on Dividend Irrelevance and the Gordon Valuation Model," *Journal of Finance*, December 1971, *26*, 1115–22.

Brickley, James A., "Shareholder Wealth, Information Signalling, and the Specially Designated Dividend: An Empirical Study," *Journal of Financial Economics*, August 1983, *12*, 187–209.

Brudney, Victor, "Dividends, Discretion, and Disclosure," *Virginia Law Review*, November 1980, *66*, 85–129.

Carlton, Dennis W. and Fischel, Daniel R., "The Regulation of Insider Trading," *Stanford Law Review*, May 1983, *35*, 857–95.

DeAngelo, Linda E., "Auditor Size and Audit Quality," *Journal of Accounting and Economics*, July 1981, *3*, 183–99.

Diamond, Douglas W. and Verrecchia, Robert E., "Optimal Managerial Contracts and Equilibrium Security Prices," *Journal of Finance*, May 1982, *37*, 275–87.

Easterbrook, Frank H., "Insider Trading, Secret Agents, Evidentiary Privileges, and the Production of Information," *Supreme Court Review, 1981*, 309–65.

______ **and Fischel, Daniel R.,** "Auctions and Sunk Costs in Tender Offers," *Stanford Law Review*, November 1982, *35*, 1–21.

______ **and** ______, (1983a) "Voting in Corporate Law," *Journal of Law and Economics*, June 1983, *26*, 395–425.

______ **and** ______, (1983b) "Mandatory Disclosure and the Protection of Investors," Working Paper No. 17, Law and Economics Program, University of Chicago, October 1983 (*Virginia Law Review*, May 1984).

Fama, Eugene F., "Agency Problems and the Theory of the Firm," *Journal of Political Economy*, April 1980, *88*, 288–307.

______ **and Jensen, Michael C.,** (1983a) "Agency Problems and Residual Claims," *Journal of Law and Economics*, June 1983, *26*, 327–49.

______ **and** ______, (1983b) "Organizational Forms and Investment Decisions," Working Paper No. MERC 83-03, Managerial Economics Research Center, University of Rochester, June 1983.

Feldstein, Martin and Green, Jerry, "Why Do Companies Pay Dividends?," *American Economic Review*, March 1983, *73*, 17–30.

Fischel, Daniel R., "The Law and Economics of Dividend Policy," *Virginia Law Review*, May 1981, *67*, 669–726.

Gordon, M. J., "Dividends, Earnings, and Stock Prices," *Review of Economics and Statistics*, May 1959, *41*, 99.

Grossman, Sanford J., "The Informational Role of Warranties and Private Disclosure About Product Quality," *Journal of Law and Economics*, December 1981, *24*, 461–83.

______ **and Hart, Oliver,** "Corporate Financial Structure and Managerial Incentives," in J. J. McCall, ed., *The Economics of*

Information and Uncertainty, Chicago: University of Chicago Press, 1982.

Hakansson, Nils H., "To Pay or Not to Pay Dividend," *Journal of Finance*, May 1982, *37*, 415–28.

Handjinicolaou, George and Kalay, Avner, "Wealth Redistribution or Informational Content: An Analysis of Returns to the Bondholders and to the Stockholders," draft, July 1982.

Hansen, Robert S. and Pinkerton, John M., "Direct Equity Financing: A Resolution of a Paradox," *Journal of Finance*, June 1982, *37*, 651–65.

Holmstrom, Bengt, "Moral Hazard in Teams," *Bell Journal of Economics*, Autumn 1982, *13*, 324–40.

Jensen, Michael C. and Meckling, William, "Theory of the Firm: Managerial Behavior, Agency Costs, and Capital Structure," *Journal of Financial Economics*, October 1976, *3*, 305–60.

John, Kose and Kalay, Avner, "Costly Contracting and Optimal Payout Constraints," *Journal of Finance*, May 1982, *37*, 457–70.

Levmore, Saul, "Monitors and Freeriders in Commercial and Corporate Settings," *Yale Law Journal*, November 1982, *91*, 49–83.

Litzenberger, Robert H. and Ramaswamy, Krishna, The Effects of Dividends on Common Stock Process: Tax Effects or Information Effects?," *Journal of Finance*, May 1982, *37*, 429–43.

Manne, Henry G., *Insider Trading and the Stock Market*, New York: Free Press, 1966.

Manning, Bayless, *Legal Capital*, St. Paul: West Publishing, 1981.

Marcus, Alan J., "Risk Sharing and the Theory of the Firm," *Bell Journal of Economics*, Autumn 1982, *13*, 369–78.

Miller, Merton H. and Modigliani, Franco, "Dividend Policy, Growth, and the Valuation of Shares," *Journal of Business*, October 1961, *34*, 411–13.

______ **and Rock, Kevin**, "Dividend Policy Under Asymmetric Information," draft, November 1982.

______ **and Scholes, Myron S.**, "Dividends and Taxes," *Journal of Financial Economics*, December 1978, *6*, 333–64.

______ **and** ______, "Dividends and Taxes: Some Empirical Evidence," *Journal of Political Economy*, December 1982, *90*, 1108–42.

Modigliani, Franco, "Debt, Dividend Policy, Taxes, Inflation and Market Valuation," *Journal of Finance*, May 1982, *37*, 255–73.

______ **and Miller, Merton H.**, "The Cost of Capital, Corporation Finance and the Theory of Investment," *American Economic Review*, June 1958, *48*, 261–97.

Penman, Stephen H., "The Predictive Content of Earnings Forecasts and Dividends," *Journal of Finance*, September 1983, *38*, 1181–99.

Ross, Steven S., "The Determination of Financial Structure: The Incentive Signalling Approach," *Bell Journal of Economics*, Spring 1977, *8*, 23–40.

Rozeff, Michael S., "Growth, Beta and Agency Costs as Determinants of Dividend Payout Ratios," *Journal of Financial Research*, Fall 1982, *5*, 249–59.

Shavell, Steven, "Risk Sharing and Incentives in the Principal and Agent Relationship," *Bell Journal of Economics*, Spring 1979, *10*, 55–73.

Shiller, Robert H., "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?," *American Economic Review*, June 1981, *71*, 421–36.

Smith, Clifford W. and Warner, Jerold B., "On Financial Contracting: An Analysis of Bond Covenants," *Journal of Financial Economics*, June 1979, *7*, 117.

Stiglitz, Joseph, "On the Irrelevance of Corporate Financial Policy," *American Economic Review*, September 1974, *64*, 851–66.

Symposium (on the Theory of the Firm, various authors), *Journal of Law and Economics*, June 1983, *26*, 237–427.

Managerial Rents and Outside Recruitment in the Coasian Firm

By ROGER L. FAITH, RICHARD S. HIGGINS, AND ROBERT D. TOLLISON*

In the corporation, decision making about the use of capital by managers and risk taking by the owners of capital are specialized activities. Various institutional mechanisms have been analyzed as forces disciplining the opportunistic behavior of management under these conditions. We maintain that no single mechanism, such as wage revision in the managerial labor market or an internal device that relies on influencing incentives by making executives residual claimants in the firm, is sufficient to solve problems of managerial incentives. Instead, the margins of several control devices are extended simultaneously to maximize the value of the firm, net of the cost of control. In this context we will emphasize one means of managerial control which has not been fully analyzed to date—the impact of outside hiring on the behavior of incumbent managers.

Our analysis is developed in terms of the neoclassical theory of the firm, as exposited by Ronald Coase (1937), and extended by Armen Alchian and Harold Demsetz (1972), Michael Jensen and William Meckling (1976), Eugene Fama (1980), and others. In this tradition the firm is seen as an institution that economizes on transaction costs. It is defined as a collection of contracts between input owners and the owner(s) of the firm where multilateral contracting among resource owners is replaced by bilateral contracting between each resource owner and the firm. This contractual structure reduces transaction costs because it reduces the number of formal contracts required for productive activity.

Alchian and Demsetz extend the theory of the firm by analyzing the substitution of hierarchical organization for market organization—the productivity of team production outweighs the cost of monitoring individual marginal value contributions. Firms deal with the monitoring problem by establishing a specialized monitoring input that is common to all input contracts and that can renegotiate individual contracts. The incentives of this specialized input are controlled by making it a residual claimant. Alchian and Demsetz appreciate the value of further specialization in the modern corporation, which requires substantial capital to take advantage of available economies of scale. The classical owner-entrepreneur is replaced by numerous owners, who hold diversified portfolios of ownership claims and who delegate decision-making authority to management. In Alchian and Demsetz, the problem of separation of ownership and control is dealt with in numerous ways. Their emphasis, however, is on the internal and external control mechanisms related to the transferability of voting shares.

The ownership-control problem is a major theme in Jensen and Meckling. They recognize that efficient allocation of risks calls for the specialization of risk-bearing and decision-making activities within the firm, and that efficient allocation is achieved at a cost of reduced control over management by the firm's owners. They emphasize various mechanisms devised in the interest of the firm's owners to control opportunistic behavior by management—contingency clauses in shareholders' contracts, the composition of man-

*Department of Economics, Arizona State University, Tempe, AZ 85287; Economist, Federal Trade Commission, Washington, D.C. 20580; and Center for Study of Public Choice, George Mason University, Fairfax, VA 22030, and Department of Economics, Clemson University, 222 Sirrine Hall, Clemson, SC 29631, respectively. The opinions expressed in this paper are solely our own and do not necessarily reflect the views of the Federal Trade Commission, its staff, or any individual Commissioner. We are grateful to James Buchanan, Mark Crain, Robert McCormick, Dennis Mueller, Roger Sherman, and William Shughart for helpful comments on earlier drafts. The usual caveat applies.

agerial pay, proxy fights, stockholders' derivative actions, and so on. They are also sensitive to the internal and external mechanisms of control based on efficient valuation of shareholder claims in the capital market, a process that also plays a paramount role in Henry Manne's (1965) takeover model of corporate control.[1]

Fama argues that the ownership-control problem can be resolved by wage revision in the managerial labor market. His thesis parallels that of Alchian and Demsetz, which also relies heavily on external markets to achieve internal managerial discipline. We agree with Fama that the degree of discipline imposed by the managerial labor market is an empirical question, but a priori, we are inclined toward an eclectic hypothesis which accommodates several margins of internal and external control.

Specifically, we present a model of the determination of a firm's outside hiring rate. Just as the firm externally substitutes hierarchical organization for market organization and internally adjusts the organizational form until returns are equalized at the margin, the firm also utilizes external and internal markets for managerial services until net gains are exhausted. Our thesis is similar to Fama's in its emphasis on the external labor market, but unlike Fama, we stress the impact of outside hiring on the matrix of internal managerial incentives, not just on the incentives of the manager(s) replaced.[2]

In Section I, we describe how managers can appropriate rents in the firm through informal (unwritten) contracts, and we suggest how such behavior can be controlled through outside hiring. In Section II, we present a formal model of the informal contracting idea. We derive implications about the relationships between firm-specific factors and the tenure and pay of the firm's chief executive officer. In Section III, we provide a test of the major implications of the model. In Section IV, we present some concluding remarks, stressing other implications of the analysis.

I. Informal Contracting

A contract is an allocation of rewards over states of the world and over actions of the contracting parties. An efficient contract will assign rewards based on joint action-state contingencies in order to provide incentives for value-maximizing actions, provided that the costs of contracting (costs associated with enumerating states and state-contingent actions, verifying states and actions, and enforcing judgments) are not prohibitive. An efficient contract maximizes the expected net gains from trade.

Because it is costly to write and enforce completely specified contracts, many contract terms are left implicit. To the extent that some terms of contracts are implicit, appropriate actions for given states of the world are not expressly stated. Instead, they are assessed *ex post*, with costs and rewards assigned accordingly.

A contract can be enforced in two major ways: either a penalty is imposed by a third-party in the event of breach, or the prospect of gain from continued dealing is withheld from the breaching party. In the case of a self-enforcing agreement, no outside party is required to detect violations, assess damages, and enforce judgments. Rather, the parties to the contract continue to abide by the agreement so long as the expected net gains exceed the current gains of violating it. Monitoring activities may still go on within self-enforcing agreements since the value of an agreement is greater the more quickly violations are detected. Parties will exhaust the marginal gains of more rapid detection through monitoring.

[1]Benjamin Klein, Robert Crawford, and Alchian (1978) provide further analysis of the putative conflicting interests in agent-principal relations. They cite the role of appropriable quasi rents in creating incentives for *ex post* opportunistic behavior, and they highlight various methods of market contract enforcement with an emphasis on vertical integration. The cost of contracting and the limits of market contract enforcement are also major themes in Lester Telser's (1980) paper on self-enforcing agreements.

[2]It is also important to note that because market participants are rational, management interests are also served by effective control devices. Rational investors will anticipate managerial discretion, and contracts governing managerial rewards will be adjusted to compensate for any anticipated exercise of expense preference by management.

Another important distinction is that between exogenous and endogenous enforcement. When contracts are multilateral, the term "third-party enforcement" is ambiguous. The mutual responsibilities of two agents in a multilateral agreement can be enforced by another contracting party. We refer to this situation as endogenous third-party enforcement to distinguish it from court enforcement, where the enforcer is outside the agreement, and from self-enforcing contracts, where no third-party enforcer is necessary. We use the term informal contract to denote any behavioral agreement among two or more managers which is self-enforcing or enforced endogenously.

As a by-product of management's putative role in monitoring team production, managers are well placed to capture a portion of the firm's value through informal contracting. Legitimate monitoring devices can be used to detect incipient and recent violations, encouraging self-enforcing agreements that serve managerial goals at the expense of the firm. The internal organization established to discipline managerial discretion provides a system of checks, which empowers managers to penalize actions that violate informal agreements, as well as to enforce the formal contracts that define the firm. For example, manager C may be in a position to penalize manager A for violating a provision of an informal agreement governing the duties of manager A and manager B. In turn, manager B may be in a position to enforce the subcontract between managers C and A, and so on. This is endogenous third-party enforcement of informal contracts as defined above.[3]

When there are potential gains from trade inside the firm and some agents are situated to control the magnitude and distribution of these gains, informal contracting will compete with other mechanisms devised to control the actions of the agents on behalf of the owner(s). These latter mechanisms—formal contracts, internal organization, and external market price adjustment—are not established in a vacuum, but evolve interdependently with the informal contracting activity of management, and the mechanism of control based on internal organization provides a means of facilitating informal agreements.

Of course, mechanisms of control will respond to informal contracting. One method of control, which has large potential for constraining informal contracting, is hiring outside managers. In this respect our emphasis is different from that of Fama. Full *ex post* "settling up" is not necessary in our model. What is important is that outsiders replace key managerial personnel, thereby upsetting the informal contracts established by incumbent managers. Our emphasis is not on replacing a manager with a more productive outsider, but on raising the net productivity of the entire management team by dissolving informal contracts.

II. Theory and Implications

We analyze the problem of control of the management team by the owner(s) of the firm through resort to the market for managers. We consider both profit-maximizing firms and firms in which ownership rights are attenuated.

A. *The Market for Managers*

Every currently employed manager is an "insider" with respect to his current employer and an "outsider" with respect to other firms. A competitive labor market implies that the marginal return to working for

[3]Suppose that A, B, and C comprise a pool of inside candidates for a position at a higher level in the firm. Rather than compete for promotion by "overworking" or informing on the rent-seeking activities of others, the candidates may collude not to destroy the rents of one's former colleagues if they are promoted. The competition to be promoted will take place in terms of informal contracts that are made with one's colleagues in return for support in the internal reviews or audits necessary for promotion. These informal contracts can take many forms—larger raises in the future, better office and support facilities, increased subsequent influence in the firm, and so on. Alternatively, managers may trade damaging inside secrets about one another's rent-seeking behavior, with a commitment to blow the whistle on any manager (including the one promoted) who seeks to use similar information to improve his future promotion probabilities.

any particular employer will be equalized. This return is made up of both the contractual wage, W, and the rents, R, accruing to the manager. Assuming a competitive labor market and using subscripts 1 and 0 to denote insiders and outsiders, we can write

$$(1) \quad W_0 + R_0 = W_1 + R_1 = MP(V),$$

where V is a vector of variables which affect managerial marginal productivity or performance.

To the firm, there is a difference in the costs of employing an inside manager for another period or promoting an insider to a higher position versus bringing in an outside manager. If the firm goes outside, it incurs a cost equal to $W_0 + R_0 + OH$, where OH is the cost of outside hiring. The firm gains $W_1 + R_1 + R_S$, where R_S represents the rents appropriated by management through informal contracts, *throughout the organization*, destroyed by bringing in an outsider.

In equilibrium, the firm will equate the marginal benefits and marginal costs of outside hiring:

$$(2) \quad W_0 + R_0 + OH = W_1 + R_1 + R_S,$$

$$\text{or} \quad W_0 - W_1 = R_1 - R_0 + R_S - OH.$$

Given (1), the firm goes outside until $R_S = OH$, but we will find it useful to concentrate on net managerial rents, $R_N = R_1 - R_0 + R_S$. Thus, given market wages, the basic cost relationship facing the firm is a tradeoff between net managerial rents R_N, and the cost of outside hiring OH.

B. *Profit-Maximizing Firms*

The firm's tradeoff is derived from a pair of cost functions relating managerial rents R_N, and the cost of outside hiring OH, to the rate of outside hiring, h.

The managerial rent function is assumed to be an increasing function of managerial tenure in the firm, T, and it is assumed that $R_0 < R_1(T)$ for all positive values of T. That is, the appropriable rents of a new manager, R_0, are less than the appropriable rents of a current inside manager with positive tenure. Since average tenure is inversely related to the outside hiring rate, managerial rent declines in h. This is because the outside candidate is not bound by and does not receive private benefits from the informal agreements that the inside managers have made. Since the marginal gain of outside hiring will eventually be dominated by losses caused by too frequent interruption of legitimate lines of communication among members of the managerial team, the marginal reductions in managerial rent appropriation decline.[4] Thus,

$$(3) \quad R_N = f(h, X), \quad f_h < 0 \quad \text{and} \quad f_{hh} > 0,$$

where X is a vector of other exogenous variables which affect R_N.

The cost of outside hiring is primarily the cost of bringing the outsider's productivity up to the level of the insider. This cost arises because insiders have been trained in firm procedures and own specific capital which is valuable to the firm. The greater the rate of outside hiring, the greater these training costs. In addition, there are costs of finding the outside manager. It is costly to detect fakers (low-productivity types) in the supply of outside candidates. It is assumed that the marginal cost of error (higher training costs to offset lower productivity) rises the more the firm resorts to the outside market. Therefore,

$$(4) \quad OH = g(h, Y), \quad g_h > 0 \quad \text{and} \quad g_{hh} > 0,$$

where Y is a vector of other exogenous variables which affect OH.

Equations (3) and (4) represent the benefits and costs of hiring on the outside. As the outside hiring rate goes up, the managerial team is subjected to stricter discipline, and managerial rent appropriation is reduced. At the same time, the firm incurs a loss in value because there is a cost of outside hiring. A profit-maximizing owner will choose an out-

[4]An alternative categorization of costs and benefits would treat the negative impact of greater outside hiring on informal contracting as an element of outside hiring costs. The destruction of the legitimate ability to communicate occasioned by management instability could be treated as depreciation of firm-specific capital.

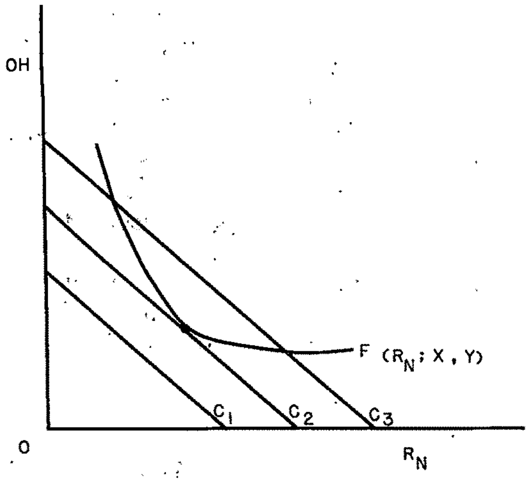


FIGURE 1

side hiring rate that minimizes the sum of R_N and OH because he is indifferent to the source of the costs that diminish the firm's value.

Analytically, the R_N-OH tradeoff is constructed by eliminating h between (3) and (4). The resulting opportunity locus is denoted $OH = F(R_N; X, Y)$, where $F' < 0$ and $F'' > 0$. The constraint F is depicted in Figure 1. The value of h is implicit in F; as one moves from right to left along F, h increases.

The profit-maximizing owner's objective is to minimize $R_N + OH$. A \$1 increase in outside hiring cost that is accompanied by a \$1 reduction in managerial rent seeking leaves the value of the firm unchanged and also leaves the owner's utility unchanged. In Figure 1, the owner's indifference lines over R_N and OH, denoted by C, have slopes equal to -1 everywhere; he is indifferent to the composition of $MR + OH$.

The actual position of the constraint F will depend upon the exogenous factors X and Y, such as the size of the management team, the level of output, the effectiveness of other managerial monitoring devices, and so on. For example, an increase in the efficiency of internal auditing would shift F toward the origin, implying a lower level of $(R_N + OH)$ at the optimal hiring rate.

The marginal condition defining the minimum cost combination of OH and R_N (given F) is

$$(5)\qquad \partial OH/\partial R_N = g_h(h^*; X, Y)/f_h(h^*; X, Y) = -1,$$

$$\text{or}\qquad g_h(h^*; X, Y) = |f_h(h^*; X, Y)|.$$

At the optimal rate of outside hiring, h^*, implicit in (5), the rate at which a dollar of managerial rent can be saved per dollar of outside hiring cost expended is equal to unity. A positive level of managerial rents and outside hiring costs at h^* imply, from (2), that the firm's contractual wage for insiders declines, relative to the outsider's wage, with increasing tenure in the firm. Further, since $R_0 < R_1[T(h^*)]$, for $h^* > 0$, it follows from (1) that $W_0 > W_1$.

C. *Derivation of the CEO Pay-Tenure Relation*

The chief executive officer (CEO) of a corporation shares in the rents created through the cooperation of the managerial team. This is the reason that the owners of the firm will sometimes resort to the outside market for a CEO. In the market for managerial services, however, total returns (rent and nonrent) are equalized across managerial positions requiring equal abilities. As discussed above, the market equilibrium relationship between nonrent pay and pay received as rent is negative, holding constant such factors as firm performance and the composition of managerial compensation.

Substituting the individual firm's equilibrium relationship between managerial rent and outside hiring, $R_N^* = f(h^*, X)$, into the equilibrium pay equation (2), we obtain a relationship between CEO nonrent pay and the variables h^* and X. Based on the qualitative information presumed about the managerial rent function ($f_h < 0$), we predict, *ceteris paribus*, that h^* impacts on W positively. An increase in reliance on the market

for managers reduces managerial rent, and lower managerial rent is associated with higher managerial nonrent pay.[5]

Two important exogenous variables, included in X, directly and indirectly (through h^*), affect *CEO* nonrent pay: the size of the management team, N, and the amount of potential rent appropriable by managers. For each of these variables, the overall impact on *CEO* nonrent pay is ambiguous, although the effects on pay are predictable a priori, holding h^* constant. In the case of the size of the managerial team, informal contracting will be more difficult the larger the number of managers whose cooperation is necessary for rent creation, that is, $\partial R_N^*/\partial N < 0$. Consequently, holding h^* constant, W is predicted to be positively related to N. The second factor is the potential rent that management can appropriate through informal contracting. An increase in the amount there is to appropriate will raise R_N^*; hence, in the *CEO* nonrent pay equation, W will be negatively related to appropriable rents, holding h^* constant.

We do not have separate measures for the size of the managerial team and the amount of appropriable rents. In our empirical test we use a measure of firm size as a proxy for these factors—on average, the greater a firm's size, the larger its managerial team and its capital in jeopardy. Additionally, if greater ability is required to manage large firms effectively, firm size may also affect *CEO* nonrent pay positively through $MP(V)$ in (1). The net impact of firm size on *CEO* pay is ambiguous.

The market equilibrium pay condition in equation (1) implies a *ceteris paribus* relationship between W and V, the vector of parameters that affect managerial marginal productivity. One element of V is the extent that management is remunerated with contingent pay. The composition of pay is important in controlling managerial behavior. It is a substitute for outside hiring to some extent because the interest of owners in maximizing the firm's value can be imposed on the manager by compensating him through stock options and bonuses. Pay composition is not a perfect substitute for other control devices, however; the manager requires a pay package with a larger expected value because diversification is restricted and risk increased. In our approach, the firm will operate on both the outside hiring and contingent pay margins. Since managerial rent seeking will be negatively related to contingent pay compensation, we predict that *CEO* nonrent pay will be positively related to the contingent pay variable.

Another component of V is firm performance. Because firm performance is an *ex post* indicator of the *CEO*'s productivity, a measure of firm performance should be positively related to nonrent pay.

D. *Attenuated Property Rights*

There are firms in which the property rights of owners are limited by law as in the case of firms subject to rate of return regulation. There are also firms and nonprofit organizations in which the transferability of property rights is limited. That managers in public utilities, labor unions, and universities engage in expense preference is not controversial. What has not been generally analyzed is the role of outside hiring in enterprises with attenuated property rights. In this section we are concerned with the effect of property rights attenuation on the organization's use of outside hiring to discipline management.[6] Specifically, we contrast the

[5] Our model also predicts that managerial pay will be greater at all levels of management as the frequency of hiring outside managers increases because expected managerial rent appropriation falls throughout the firm. Further, lower-level managers will revise their *ex ante* expectations regarding promotion probabilities as the observed outside hiring rate increases. In order to prevent more than the optimal rate of voluntary turnover (quits), the firm will have to compensate existing and entering managers for the increased risk of slower promotion. However, the firm alternatively could compensate lower-level managers by reducing the size of internal candidate pools, thereby maintaining the probability of promotion and expected utility.

[6] Attenuating property rights may have the effect of reducing the demand and supply of internal and exter-

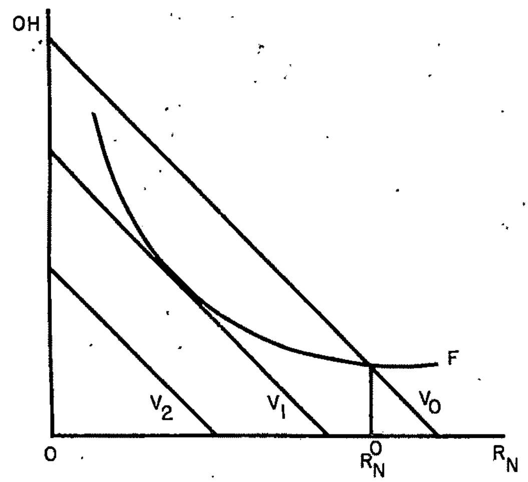


FIGURE 2

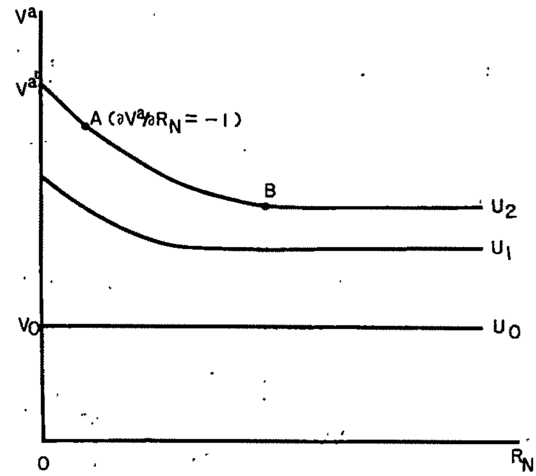


FIGURE 3

CEO pay relation and *CEO* turnover policy in firms with and without rate of return regulation.

In firms subject to a rate of return constraint, management is prohibited from maximizing firm value. Additionally, a profit restraint can alter input prices. We focus on the impact of a rate of return constraint on the value to management of managerial rents in terms of the firm's profit, which affects the firm's tenure policy. For example, in Figure 2, if the rate of return constraint limited the value of the firm (the owners' maximand) to V_0, which is below its maximum value V_1, outside hiring to discipline managerial rent seeking would be redundant for R_N less than R_N^0. In effect, the *CEO*'s indifference curves in Figure 2 would be vertical for all points below V_0. Although this illustration conveys our basic result—that the relative value of outside hiring as a means of managerial control is lower for regulated firms, resulting in greater *CEO* tenure—it is unrealistic in supposing that there is some maximum value of the firm subject to a regulatory restraint that is independent of the firm's turnover policy. Below we construct a model of managerial rent seeking in which regulation induces expense preference according to a neoclassical utility function and in which firm value is endogenous.

We suppose, in general, that the firm maximizes a utility function with arguments R_N and the actual value of the firm, V^a: $U(R_N, V^a)$. We further suppose that $V^a = V^p - (R_N + OH)$, where V^p is the potential value of the firm. For a profit-maximizing firm, $\partial U/\partial R_N \equiv 0$, and maximizing U is equivalent to maximizing V^a or minimizing $R_N + OH$. But for a regulated firm, $\partial U/\partial R_N$ is not identically equal to zero. For firms under a regulatory constraint, we assume that the indifference curves have the form illustrated in Figure 3.

In Figure 3 there is a threshold value of the firm under a regulatory constraint, V_0, below which the *CEO*'s marginal valuation of rents in terms of profits is zero. In effect, the probability that the *CEO* will be fired (on account of disciplinary action by stockholders or other firms) is one. For firm values greater than V_0, disciplinary action is less certain. Specifically, we assume that disciplinary action is less likely the greater is V^a relative to V_0, *ceteris paribus*, and given V^a, the lower is R_N. Thus, for $V^a > V_0$, the *CEO*'s marginal valuation of rents in terms

nal mechanisms for controlling managerial rent seeking. Moreover, the various control mechanisms may be affected differentially. In other words, there are "wealth" and "substitution" effects associated with attenuating property rights. In our model, we focus on the wealth effect; we analyze the effects of regulation as if the value of all control mechanisms were reduced.

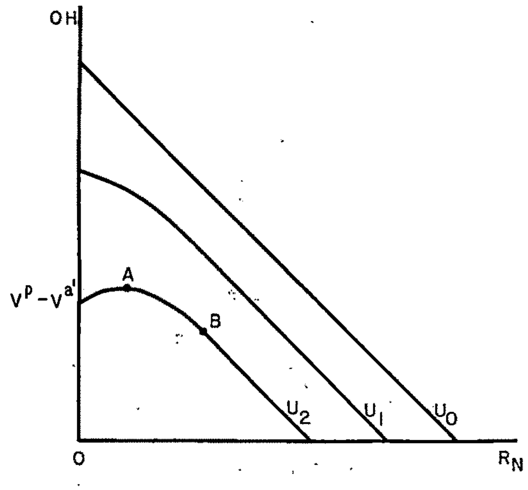


FIGURE 4

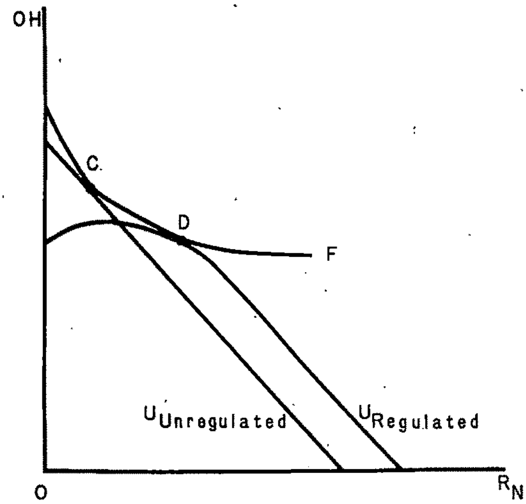


FIGURE 5

of profit rises as V^a increases, and the marginal valuation falls as R_N rises, *ceteris paribus*. These assumptions imply smooth substitution of rents for profit with diminishing marginal valuation exhibited along an indifference curve up to some value of R_N.

Given a utility function of this form, we next derive indifference curves in R_N-OH space in order to contrast the turnover policies of unregulated and regulated firms. Solving $U[R_N, V^p - (R_N + OH)] = U^0$ for OH as a function of R_N, we find that $\partial OH/\partial R_N = (\partial U/\partial R_N - \partial U/\partial V^a)/(\partial U/\partial V^a)$. Based on the second-order properties of U illustrated in Figure 3, we depict the preference map over R_N-OH in Figure 4. In contrast to Figure 3, utility decreases towards the northeast in Figure 4.

The preference map in Figure 4 is found by choosing a point in R_N-V^a space of Figure 3 and mapping it to the point $(R_N, V^p - V^a - R_N)$ in Figure 4, with slope equal to $\partial OH/\partial R_N = (\partial U/\partial R_N/\partial U/\partial V^a) - 1$. Thus, all the points on the flat indifference curve, U_0, in Figure 3 are mapped to U_0 in Figure 4, where $\partial OH/\partial R_N = -1$. Similarly, the points on U_2 in Figure 3 map to U_2 in Figure 4. Specifically, the point $(0, V^{a'})$ in Figure 3 maps to $(0, V^p - V^{a'})$ in Figure 4, with slope equal to $(\partial U/\partial R_N/\partial U/\partial V^a) - 1 > 0$. As R_N is substituted for V^a along U_2 in Figure 3, $(\partial U/\partial R_N/\partial U/\partial V^a)$ approaches unity so that point A in Figure 3 corresponds to point A in Figure 4 where the slope of the indifference curve is $1 - 1 = 0$. Beyond point B along U_2 in Figure 3, the slope of the indifference curve for the corresponding points in Figure 4 equals -1.

With the regulated-firm *CEO*'s preferences defined on R_N-OH space in Figure 4, we can readily compare the unregulated firm's choice of turnover policy and managerial rents with the regulated firm's decision. (Recall that for the unregulated firm the indifference curves in R_N-OH space are all linear with a slope of -1.) In Figure 5 we depict the respective solutions, C and D, to the managerial control problem for an unregulated and a regulated firm. The sum of R_N and OH is greater (since we know that $R_N + OH$ is minimized at C) at D than at C. Furthermore, we see that R_N is larger at D than at C; the firm with attenuated property rights chooses an outside hiring policy that allows more rents to be captured internally by management. And, based on the derivation of the constraint F, we know that the outside hiring rate is lower at D than at C. Thus, if we distinguish firms by presence of rate of return regulation, we expect to find lower outside hiring rates where ownership rights are attenuated.

The model has implications for managerial pay as well. Competition in the market for

managers will assure that managers in profit-restrained firms receive lower total pay, excluding rents, than managers in unrestrained firms where rents are lower. In a pay equation with *CEO* tenure, pay composition, firm size, firm performance, and a dummy variable indicating the presence of rate of return regulation as arguments, we predict that tenure will affect pay negatively. Moreover, a negative relationship between regulation and pay is consistent with our argument that all margins of managerial control are worth less in a regulated environment. We discuss this point in more detail below.

III. Empirical Results

A. *Regression Model*

We estimated the parameters of the following *CEO* pay equation to test the predictions of our model about tenure and regulation:

$$(6) \quad CEOPAY = a_0 + a_1 CTGPAY + a_2 COYRS + a_3 CEOYRS + a_4 (CEOYRS)^2 + a_5 SIZE + a_6 AGE + a_7 FPERF + a_8 D + u.$$

The variables in (6) are defined as follows:

CEOPAY is *CEO* compensation;

CTGPAY is contingent remuneration plus stock options divided by total compensation;

COYRS is the number of years a *CEO* has been with a company;

CEOYRS is the number of years the *CEO* has been a company's *CEO*;

SIZE is firm size measured by net sales;

AGE is *CEO* age;

FPERF is a measure of short-run firm performance;

D is a dummy variable for the presence of rate of return regulation which equals 1 if regulated and 0 otherwise; and

u is a stochastic error term with the usual properties.

This empirical model of *CEOPAY* yields several *ceteris paribus* predictions. The effect of tenure in the company on *CEOPAY* is paramount. We predict a negative sign on *COYRS* because outside hiring frequency (the inverse of tenure) disciplines informal contracting and because nonrent compensation is inversely related to pay as rent.

Second, we predict a positive sign for *CEOYRS* as an indication of a learning effect among *CEO*s. However, this argument interacts with our tenure prediction. As *CEOYRS* increases, there is both learning and reduced managerial discipline with respect to informal contracting. To capture these effects we include both a linear and quadratic term for *CEOYRS*. We predict a positive linear (learning) and a negative quadratic (as informal contracting becomes more pronounced) effect.

Third, our model predicts less *CEOPAY* in regulated firms. However, this prediction is not a *ceteris paribus* prediction because *CEOPAY* is affected by regulation through firms' tenure policies. Our model is silent about the direct effects of regulation on *CEOPAY*, holding tenure constant, although a negative regulation effect is consistent with our basic argument. The essence of our analysis is that the firm operates on several margins to control informal contracting by management. Since all margins of owner control are worth less in a regulated environment, and since we are unable to control for all of these disciplinary margins in our empirical specification, it is likely that the regulation variable will have some separate negative impact on *CEOPAY*.

Fourth, we can make a priori predictions about the impact of the remaining control variables. *CTGPAY* should impact on *CEOPAY* positively because there is less incentive for informal contracting the more an executive is compensated on the basis of short-run performance. To the extent that sales are positively related to the size of the management hierarchy, which retards rent seeking, the predicted effect of *SIZE* is positive. Also, because more managerial ability is required in large firms to achieve equal performance between small and large firms, *SIZE* will be positively related to *CEOPAY*. To the extent that sales are positively related to the firm's appropriable capital, *SIZE* will be negatively related to *CEOPAY*. Thus, the net impact of *SIZE* on *CEOPAY* is uncertain. Further, we expect a positive sign on *AGE* as a reflection of nonspecific human

capital in *CEO*s. Finally, *FPERF*, a measure of the recent performance of the *CEO*, should impact on *CEOPAY* positively.

B. *The Data*

We estimated the regression model using data on executive pay and tenure from a survey of executive compensation published by *Forbes* (1981) and on firm size from *Compustat* (1980). The sample from *Forbes* included 818 U.S. firms chosen on the basis of magnitude of total compensation; we were able to match firm data from *Compustat* for 814 firms.

We estimated several specifications of (6), depending on how much *CEOPAY* reflected contingent pay components. The *Forbes* survey separated total compensation into four categories: salary and bonuses, benefits, contingent remuneration, and stock options.[7] The data on *COYRS*, *CEOYRS*, and *AGE* were taken directly from *Forbes*. As a measure of firm size, we used the net sales figure from *Compustat*. Firm performance (*FPERF*) was measured by the average of the rates of return on equity for 1978, 1979, and 1980. More specifically, we calculated annual rates of return based on beginning-of-the year and end-of-the-year stock prices and on dividends paid during the year, and we averaged these rates over the three years prior to and including the year for which we had observations on *CEOPAY*. Finally, regulated firms were defined as public utilities. There were fifty-two public utilities in the sample.

C. *The Results*

For each of the alternative measures of *CEOPAY*, we estimated regression models with a complete set of regulation-interaction terms. None of the interaction terms was statistically significant. We subsequently dropped the interaction terms and computed the following least squares estimates.[8]

$$\text{(7a)}\quad \begin{aligned} CEOPAY1 = {} & \underset{(-7.85)}{-459.84} + \underset{(3.14)}{6.49}CEOYRS \\ & \underset{(-2.46)}{-0.17}(CEOYRS)^2 - \underset{(-4.13)}{2.17}COYRS \\ & + \underset{(22.74)}{107.35}SIZE + \underset{(1.39)}{1.31}AGE + \underset{(1.75)}{50.51}CTGPAY \\ & + \underset{(4.67)}{87.24}FPERF - \underset{(-6.42)}{138.41}D \end{aligned}$$

$R^2 = .45$; F-statistic $= 77.9$; $N = 814$.

$$\text{(7b)}\quad \begin{aligned} CEOPAY2 = {} & \underset{(-7.56)}{-518.54} + \underset{(3.49)}{8.43}CEOYRS \\ & \underset{(-2.77)}{-0.23}(CEOYRS)^2 - \underset{(-4.10)}{2.53}COYRS \\ & + \underset{(21.07)}{116.39}SIZE + \underset{(1.31)}{1.45}AGE + \underset{(2.52)}{85.13}CTGPAY \\ & + \underset{(4.73)}{103.32}FPERF - \underset{(-5.68)}{143.14}D \end{aligned}$$

$R^2 = .42$; F-statistic $= 68.4$; $N = 814$.

$$\text{(7c)}\quad \begin{aligned} CEOPAY3 = {} & \underset{(-7.52)}{-673.61} + \underset{(3.28)}{10.37}CEOYRS \\ & \underset{(-2.62)}{-0.28}(CEOYRS)^2 - \underset{(-4.02)}{3.24}COYRS \\ & + \underset{(18.43)}{132.99}SIZE + \underset{(1.68)}{2.43}AGE + \underset{(9.64)}{426.49}CTGPAY \\ & + \underset{(3.11)}{88.78}FPERF - \underset{(-4.64)}{152.95}D \end{aligned}$$

$R^2 = .42$; F-statistic $= 69.8$; $N = 814$.

[7]Benefits include imputed monetary values for items such as club dues, company-paid life insurance premiums, and the value of past stock bonus awards which vested during the year. Contingent remuneration "includes the amounts expensed for deferred compensation agreements...under performance related long-term incentive plans" (*Forbes*, p. 114).

[8]*CEOPAY1* is salary and bonuses; *CEOPAY2* is *CEOPAY1* plus benefits; *CEOPAY3* is *CEOPAY2* plus contingent remuneration. We also estimated the regression coefficients in an equation for *CEOPAY* that included stock options and achieved results that are not very surprising or interesting. In this specification the *SIZE* and *CTGPAY* variables explained almost all of the variation in *CEOPAY*, and the significance of the remaining coefficients was drastically reduced. The primary reason for this result is that the stock options component comprises such a large percentage of total pay that *CEOPAY3* plus stock options is naturally correlated with *CTGPAY*. Furthermore, on a priori grounds we are uninterested in explaining variation in total compensation where the latter variable includes the realized value of stock options and not their price (i.e., their value *ex ante*).

The first remarkable result is that the regressions uniformly explain a substantial amount of the variation in *CEOPAY*, and each of the regression coefficients is usually highly significant.

The major prediction of our model that *CEO* pay and company tenure are inversely related is corroborated. *Ceteris paribus*, and specifically holding *CEOYRS* constant, each year of company tenure reduces *CEOPAY* by \$2,170, \$2,530, and \$3,240, respectively. This is the central result for the informal contracting hypothesis.

The variables *CEOYRS* and $(CEOYRS)^2$ have the predicted signs, with estimated crossover points at 12.7, 12.8, and 12.7 years.[9] This does not mean that informal contracting and rent capture do not occur until 12.7 years of *CEO* tenure. It means that, on average, it takes 12-plus years for the effects of informal contracting to overcome the effects of experience on the various forms of *CEO* pay.

The predictions about the effects of *CEO* experience (*AGE*), firm size (*SIZE*), and firm performance (*FPERF*) are also corroborated. In the case of size, we used two specifications, net sales and the natural logarithm of net sales. Assuming that changes in *CEOPAY* are proportional to percentage changes in firm sales has substantially more explanatory power than assuming sales has a linear effect on pay. The results reported in (7a), (7b), and (7c) are based on the logarithm of sales. The R^2 is as much as 50 percent lower when the linear specification is adopted. Firm performance as measured by the three-year average of annual rates of return on equity has a positive and statistically significant effect on *CEOPAY*. We also experimented with a model specification in which the three annual rates of return were jointly included as separate regressors. Typically, all three were not statistically significant, but their overall contribution was about the same as the three-year average which we have reported.

We find that *CTGPAY*, which is defined as the ratio of contingent pay and stock options exercised to total *CEO* compensation, is significant in all three specifications. There is little surprise in discovering the substantial significance of *CTGPAY* in (7c), but apparently, judging by its coefficient size and significance in (7a) and (7b), contingent pay is heavily relied on to discipline managerial behavior.

The estimated coefficients of D reveal that regulation has a negative effect on *CEOPAY* which is independent of tenure and contingent pay. We interpret this to mean that there are several other control margins besides tenure and pay composition, the value of which is reduced by rate of return regulation. Furthermore, the magnitude of coefficients of D are noteworthy; *ceteris paribus*, managers in rate of return regulated firms receive on average approximately \$150,000 less pay than their counterparts in unregulated corporations.

IV. Conclusion

Our model implies several relationships which are different from theories of corporate control articulated by Oliver Williamson (1975) and Williamson, Michael Wachter, and Jeffrey Harris (1975) and based on the concept of an internal labor market. First, the internal labor market theory suggests that in order to economize on the production of costly firm-specific information, firms will provide fewer ports of entry to outside managers in the higher eschelons of the management pyramid. While it is conceded that long periods of exposure in the office environment is an ideal way to gather valuable firm-specific information, it is also a fertile breeding ground for intermanager agreements which need not serve the interests of the firm. Our theory suggests that although hiring managers from the outside

[9] The linear component of the marginal effect of *CEOYRS* on *CEOPAY* is not given simply by the coefficient on *CEOYRS*. The reason is that *COYRS* includes both company tenure as non-*CEO* and *CEO*. Thus, strictly speaking, we must add the coefficients of *CEOYRS* and *COYRS* to get the appropriate linear component. For example, to find the crossover point in (7a), we must solve the marginal condition, $2\times(0.17)CEOYRS = 6.49 - 2.17$, for *CEOYRS*.

entails costs, the gains to hiring outside managers rises as one moves up the management hierarchy.

Second, our model implies that across firms at a fixed level in the management hierarchy, managerial compensation is inversely related to tenure. This proposition is contrary to that suggested by the internal labor market theory.

Third, our model implies that property rights, including rights of market access, affect the demand and supply of monitoring and control. In other words, differences in property rights may have a wealth or substitution effect, or both. For example, rate of return regulation reduces the value to shareholders of all methods of controlling managerial rent seeking without differentially affecting the cost of any one method. Thus, our model predicts that all control margins will be contracted in a regulated environment, including *CEO* tenure. As a consequence, managerial compensation will be lower in firms subject to rate of return regulation.

Ownership differences may also affect the costs of monitoring and control as well as the gains. Some margins of control will be differentially affected by ownership form. Consider, for example, the use of hiring policy and performance-based remuneration of management. It is easier to judge the performance of a firm when shareholders can transfer their shares than when shares are nontransferable. Consequently, judging the performance of managers will be more costly in a publicly owned than a privately owned firm. Our model predicts both a wealth and a substitution effect in this instance. For public firms, all margins of control are less valuable, but at the same time controlling management by hiring more frequently from the outside will be easier than controlling management with incentive pay schemes.

Although our empirical test is confined to assessing the effects of tenure on executive pay for regulated and unregulated U.S. corporations, we think our model has wider applicability. The competitiveness of the "markets" that affect organizational performance is linked to both the value and the cost of controlling managerial rent seeking. For example, business firms in Japan are noted for their paternalistic policies toward employees, and there is no highly developed external managerial labor market. Because of the high cost of utilizing an outside hiring policy to control managerial rent seeking, we observe the predictable use of substitutes. Compared to experience in U.S. corporations, executives in Japanese firms wind their way through *all* the firm's divisions as they climb up in the firm. Beside acquiring knowledge about the firm's operations, this policy assures that would-be *CEO*s have less opportunity and incentive to invest in informal contracts within a given division or functional part of the firm.

Our model also has implications for the behavior of nonprofit bureaucracies. The military, for example, has a policy of moving personnel around frequently. In this way opportunities for internal rent seeking are reduced by limiting the potential development of friendship capital with subordinates. At first this appears to contradict our model and to support the internal labor market theory. However, our model predicts that in public organizations, both the value and cost of controlling rent seeking are affected. In some cases, the effect of public ownership is to raise the relative productivity of tenure as a means of control, even as the value of all methods of control is reduced.

Lest the reader infer that our model is tautological, we observe that in the case of the military, our model predicts that during wartime, when output is more easily monitored, there will be less movement of managerial personnel. Similarly, Christian church organizations vary between totalitarian centralism and individualistic atomism. In the Roman Catholic Church, there is more cross subsidization than there is in the Baptist Church, and hence monitoring of clerical rent seeking by parishioners is costly. Rent seeking in hierarchial churches is controlled to a large extent by planned turnover of the local pastors. Within Congregational Churches, rent seeking is controlled more through tenure based on church-specific performance evaluation.

REFERENCES

Alchian, Armen A. and Demsetz, Harold, "Production, Information Costs and Economic Organization," *American Economic Review*, December 1972, *62*, 777–95.

Coase, Ronald H., "The Nature of the Firm," *Economica*, November 1937, *4*, 386–405.

Fama, Eugene F., "Agency Problems and the Theory of the Firm," *Journal of Political Economy*, April 1980, *88*, 288–307.

Jensen, Michael C. and Meckling, William H., "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," *Journal of Financial Economics*, October 1976, *3*, 305–60.

Klein, Benjamin, Crawford, Robert G. and Alchian, Armen A., "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," *Journal of Law and Economics*, October 1978, *21*, 297–326.

Manne, Henry G., "Mergers and the Market for Corporate Control," *Journal of Political Economy*, April 1965, *73*, 110–20.

Telser, Lester G., "A Theory of Self-Enforcing Agreements," *Journal of Business*, January 1980, *53*, 27–44.

Williamson, O. E., *Markets and Hierarchies: Analysis and Antitrust Implications*, New York: The Free Press, 1975.

______, **Wachter, M. C., and Harris, J. E.,** "Understanding the Employment Relation: The Analysis of Idiosyncratic Exchange," *Bell Journal of Economics*, Spring 1975, *6*, 250–78.

Forbes, June 8, 1981.

Standard and Poor's Compustat II Services, Inc., *Compustat*, 1980.

An Empirical Analysis of Welfare Dependence

By MARK W. PLANT*

Recipients of the benefits of social welfare programs often receive benefits for prolonged periods. Some observers take this fact as evidence that we have a two-class economic system with some members continuously in need of assistance.[1] Others contend that, like a narcotic, social welfare programs create their own dependence. Once potential recipients have benefitted from a social program, it is asserted, they are less likely to become nonrecipients. Both these theories explain persistent participation in welfare programs, but the policy implications of each are very different. If our society consists of two classes, attempts by the government to decrease the length of dependence on welfare by changing the economic nature of the welfare programs will meet with little success in the short run since the structure of earnings is not sensitive to these programs. Essentially this view is one of poverty being a structural phenomenon: the poor will always be with us, in the short run at least, no matter what the configuration of the welfare system is. Alternatively, if welfare is addictive and induces recipients into dependence, then a reconfiguration of the welfare system could shorten the average length of dependence quickly. If the government increases the opportunity cost of leisure, thereby making the "welfare habit" expensive, it will shorten the period of dependence.

This paper is an attempt to sort out the empirical issues involved in distinguishing these two explanations of welfare dependence. Using data from the Seattle-Denver Income Maintenance Experiments (SIME/DIME), I explore the causes of the serial correlation in welfare dependence by formulating an econometric model of participation behavior that will allow us to distinguish among various phenomena that may result in persistent welfare participation.

I. Theoretical Model

The fact that a family's earnings are correlated over time implies that any observable variable determined by earnings will be serially correlated. In particular, since the receipt of benefits from any type of welfare program depends on earnings, welfare participation patterns in the eligible population should exhibit some serial correlation, and the extent of such correlation depends on the stochastic nature of earnings. If the welfare program also had an affect on families' labor force behavior, then the observed distribution of earnings in the population would be different in the presence of a welfare program, and the extent of serial correlation in welfare participation would depend on the economic incentives imbedded in the program.[2]

Suppose we were in a world where there were no welfare programs, and we were interested in giving financial aid to some segment of the population. In particular, suppose we wanted to help those with income below some poverty level Y_P and we could characterize the distribution of income by the density function $f(Y)$. The fraction of families who would be expected to receive aid from this program would be $F(Y_P)$, where $F(\ )$ denotes the cumulative distribution function corresponding to $f(\ )$. We could easily extend this simple exercise to a mul-

*University of California, Los Angeles, CA 90024.
This paper was taken from research done for my dissertation at Princeton University. I am grateful to Orley Ashenfelter, James Brown, George Cave, Michael Murray, and Richard Quandt for useful comments and suggestions. Financial support was provided by the National Science Foundation, the Industrial Relations Section of Princeton University, and the Institute of Industrial Relations at UCLA.

[1]For example, see Michael Harrington (1962).

[2]Several other papers have considered participation patterns and response to welfare programs over time with substantially different models. For an example using the same data as the present paper see Philip Robins and Richard West (1980a, b).

tiperiod framework, by denoting the multivariate density function for income as $f(Y_1, Y_2, \ldots, Y_n)$. The fraction of families expected to participate in all n periods would be $F(Y_P, Y_P, \ldots, Y_P)$, assuming the poverty level income remained constant over time. The fraction of families following any particular pattern of benefit receipt over time could be predicted by evaluating the multivariate distribution function appropriately.[3] These predictions depend on two assumptions. First, we must be able to characterize the multivariate distribution of income in the absence of the welfare program. Second, if the predictions are to be of any value, the existence of the welfare program must have no effect on labor force behavior or on the distribution of earnings. For example, people whose incomes are just over the poverty level Y_P must not change their behavior so as to be eligible for the income supplements. To the extent that either assumption is unwarranted, our predictions of the number of families participating in the welfare programs will be incorrect. The econometric method that I spell out takes advantage of this potential disparity to analyze the available data.

Data in the SIME/DIME were gathered on two types of families—those eligible to receive income supplements from the experimental welfare programs and those whose behavior was simply observed over time. Following the terminology of the experimental sciences, I will call the first set *experimental* families and the second, *control* families. The control families allow us to characterize the income distribution in the absence of the experimental welfare program.[4] Using this income distribution, predictions can be made about the patterns of benefit receipt among experimental families. If the programs are having no economic effect, these predictions should be very close to the actual behavior observed in the experimental sample. If, however, experimental families are reacting to the presence of the program by changing their labor force behavior and thus changing the observed income distribution, then differences will be observed between the actual and predicted behavior. The task before us as econometricians is to devise a simple method of detecting these differences and to interpret them in a meaningful way. To do this we must briefly review the nature of the welfare program offered to experimental families in SIME/DIME.

The benefit scheme used in the experiments was what is commonly referred to as a negative income tax (*NIT*) program. Benefits were calculated using the formula:

$$Benefit = G - tY,$$

where G is the guaranteed level of payments if earnings (Y) were zero, and t is referred to as the marginal tax rate.[5] The benefit paid a family was positive if

$$(1) \qquad Y < G/t.$$

The quantity G/t is referred to as the break-even level. Thus if there were no economic behavior (i.e., if Y were not a function of G or t), families receiving benefits would simply be those with income less than G/t. In fact, earnings are a function of G and t. In particular, in a simple model of labor supply, G is equivalent to an increment in lump sum income which will decrease

[3]For example, suppose the pattern in question was no-benefit receipt in the first year, benefit receipt in the second, and no-benefit receipt in the third. If a family received benefits when their income was less than Y_P, then the fraction of families expected to follow that pattern would be

$$\int_{Y_P}^{\infty}\int_{0}^{Y_P}\int_{Y_P}^{\infty} f(y_1 y_2 y_3)\, dy_1\, dy_2\, dy_3.$$

[4]This of course assumes the control families and the experimental families were randomly chosen from the same distribution. This assumption is problematic in the SIME/DIME data. For further information, see fn. 9.

[5]In the SIME/DIME, there were eleven different *NIT* programs offered to various segments of the experimental sample. These programs varied in both the choice of the guarantee level G, and the tax rate t. The level G varied between \$3,800 and \$5,600 and t between 0.5 and 0.8. In some of the programs a declining tax rate was used, where the rate of decline was 0.025 per thousand dollars of income. In my treatment of the data, for the families in the declining rate experiments I use the marginal tax rate at pre-experimental income, which was calculated by *SRI*.

hours of work and thus decrease earnings. The implicit tax on earnings, t, lowers the net wage and would result in fewer hours worked and thus less earnings.[6] Intuitively, one would expect to see eligible families whose income in the absence of the *NIT* would have been below the break-even level working less. More importantly, however, some families whose earnings in the absence of the *NIT* program would have exceeded G/t might decrease their labor supply in order to become eligible for the *NIT* benefits. If income and leisure were good substitutes, families who initially would not qualify for benefits would change their behavior so they would receive them.

Samuel Rea (1977) and Orley Ashenfelter (1983) have characterized this behavior mathematically. Using a second-order approximation to the expenditure function, they show that a family will receive benefits if and only if

$$\ln Y < \ln(G/t) + 0.5et, \tag{2}$$

where e denotes the compensated elasticity of labor supply and $\ln(x)$ denotes the natural logarithm of x. Although this a local result it gives us some foundation for examining discrete changes. The essence of equation (2) is that a family will participate if the log of income in the absence of the program were less than the log of the break-even level, plus a term that increases as leisure and income become better substitutes. The more easily income can be substituted for leisure, the larger is e, and the larger this critical earnings level. The magnitude of the effect is a positive function of the tax rate. Throughout the remainder of the paper, I use e as the summary measure of the "taste for leisure." There are two implicit assumptions to be made in characterizing the utility function by this elasticity. First, assume e is constant over all earnings levels for an individual family, and second, assume e is constant across families. Both assumptions are somewhat unrealistic. The first implies I am using a local result globally, and the second implies all families have the same shaped indifference curves, although the location may vary.[7] Any estimate of e then must be interpreted cautiously. I am attempting to characterize families' sensitivity to the tax rate which must depend on preferences. Large tax rate effects imply considerable substitution between income and leisure. Therefore the empirical measure of these tax effects I call e is at least related to the true compensated elasticity of labor supply.

This simple model of participation behavior would yield a prediction of the expected number of families receiving aid in any given year, if we knew the distribution of log-income in that period and the parameter e. However, this model alone does not incorporate the notion that the welfare system might generate its own dependence. That is, the fraction of families receiving aid may grow over time even if preferences were constant and the distribution of income were stationary. This notion of welfare addiction is what is casually referred to as the welfare trap. In statistical terms, the welfare trap hypothesis posits that the probability of any family receiving a subsidy depends on its past history of subsidy receipts. In particular, previous participation should increase the probability of current participation, *ceteris paribus*. Some analysts would be tempted to say that once a family receives an initial subsidy from a welfare program, its preferences "shift towards leisure." Such changes in preferences can explain almost any behavior, and it would be useful to characterize the increased dependence on welfare as a result of some change in the constraints on the economic actors. In my study (1982), I show that the welfare trap can arise when there are fixed costs associated with the initial receipt of the benefits. Alternatively, there may be fixed costs of leaving the welfare program and re-entering the workforce. The presence of either type of

[6] The income and substitution effects work in opposite directions, but in an *NIT* the next effect is unambiguously to decrease hours of work. See Robins and West (1980a, b).

[7] In my dissertation, the second assumption is relaxed by parameterizing e as a function of family-specific characteristics.

fixed cost means that once a family is in one state or another, it tends to stay unless earnings shift so dramatically that it is worth incurring the fixed costs to change states. This notion of state dependence is precisely what is meant by the term "welfare trap." Formally, if we let d_{s-1} equal 1 if the family is a previous participant and 0 if not, the family will receive an *NIT* payment in period s if

$$\ln Y_s < \ln(G/t) + 0.5et + (d_{s-1} - 1)(F/G), \tag{3}$$

where F denotes the monetary fixed cost of movement between states.[8] Equation (3) demonstrates that a family who has previously participated is more likely to participate than a previous nonparticipant since the critical earnings level at which a family will be eligible for payments increases by (F/G) once a family participates. Fixed costs are only one rationalization for state dependence. In this paper I do not intend to distinguish the reason for state dependence; I only wish to detect whether it is in the data.

There are three reasons for observing persistent participation over time. First, if income is highly serially correlated, then one would expect to see persistent participation simply because if equation (1) is satisfied in one period, it is likely it will be satisfied in the next period. This is what might be termed the structural or long-term reason for persistence. If the policymaker wanted to "cure" such long-term persistence, he would want to change the distribution of earnings in the population. Although decreasing the break-even level of the *NIT* program would decrease the number of participants, it would not change the pattern of participation for those who would still qualify for payments. Second, families are induced into participating in the *NIT* because of the program's generosity (those whose incomes would have been above the break-even level). Changing the parameters of the program (in particular, reducing the marginal tax rate), would reduce the number of participants, but would not alter the pattern of participation for those who still qualify. If there were substantial substitution effects, then adjustment of the tax rate leaving the break-even level fixed would decrease persistent participation. Third, if there were a welfare trap, the policymaker who was interested in decreasing persistent welfare dependence would want to force participants to move out of the welfare system by either decreasing the fixed costs of that movement or forcing the families to incur that fixed cost. It is of interest to discover the reason for persistent welfare participation, because different causes lead to different prescriptions to cure the malady.

My econometric approach is simple. Using the participation "rules" described in (1)–(3) above, the participation behavior one would expect to see in the experimental programs can be predicted using the observed income data on the control families if there were no economic behavior. I then contrast the observed participation behavior of the experimental families to that of the controls and attempt to discover to what extent the observed data can be characterized by either (2) or (3). My method yields estimates of the behavioral parameters e and F/G. The next section is devoted to this task.

II. Empirical Strategy and Results

The economic theory of participation behavior presented here lends itself to a very simple estimation procedure which allows us to gauge the extent of the behavioral effects of the *NIT*. In this section I develop that estimation method as well as an alternative maximum likelihood scheme. The results of both procedures are presented.

To predict the fraction of families that will participate in an *NIT* program, we must have some information on the income distribution. In particular, if the cumulative income distribution in any given year in the absence of the *NIT* were characterized by $F(\cdot)$, then we would want to evaluate $F(G/t)$. The *NIT* experiments were in theory designed to give us the required distri-

[8] See my dissertation for a proof of (3). This result follows Ashenfelter (1983) closely.

bution information.[9] The data on the control families provide the statistical benchmark against which the behavior of the experimental families can be measured. Assume that both the experimental and control families were drawn from the same population. If there were no change in the income distribution induced by the *NIT*, unbiased and consistent estimates of the fraction of experimental families who would receive *NIT* subsidies could be calculated using the data on the control families, by computing the percentage of control families whose income is below the break-even income level. If this simple model of participation behavior is adequate, then predictions should be close to the actual realizations. If there is a behavioral reaction to the *NIT*, the realizations will not square with the predictions. The appealing feature of this comparison is that no explicit mathematical specification of the distribution $f(\cdot)$ needs to be made. The only maintained hypothesis is that the control and experimental families are chosen from the same population.

The comparison between predicted and actual behavior for the SIME/DIME is presented in Table 1. The sample includes 958 white families with two parents present. Of these families, 591 were exposed to a *NIT* program for three years and the remaining 367 were observed as a control group.[10] The table lists the eight possible participation patterns in the three years of the experiment. It also shows the distribution of predicted and actual participation patterns for each of the 11 different *NIT* programs to which an experimental family could be assigned.[11] For example, for the break-even income level of $5,430, using the control data, 58.7 percent of the sample would be expected not to participate in any year (shown in parentheses), and 66.7 percent of the experimental families actually followed that pattern. The number of experimental families in the program is given along with a χ^2 goodness of fit statistic. The χ^2 statistic provides a test of the null hypothesis that the observed and predicted distributions were drawn from the same underlying multinomial distribution of participation patterns.[12] For example, based on the 30 experimental families exposed to the *NIT* program with a break-even income

[9]In fact, selection of experimental and control families was not random. The results in this paper adjust for this nonrandom assignment which was based on "normal income." The adjustment method varies depending on the estimation method and is explained in ensuing footnotes. For a detailed description of the assignment process see Michael Keeley and Robins (1980).

[10]Although more families were enrolled initially, I have restricted the sample in order to ensure a complete earnings history over four years. Some families were deleted from the data set because their pre-experimental income exceeded the established income truncation points.

[11]In any given year, experimental families were said to have participated in the *NIT* program if the payment made to them in either the first six months or the last six months of the year exceeded $120, since each family was paid $20 per month just for reporting their income. Experimental families who left the program were dubbed nonparticipants from the time they left. The implicit assumption is that families who left the program did so because it was not beneficial to continue their eligibility for payments. It is important to note that the predicted percentages were computed by a somewhat more complicated procedure than would be obvious from the text. Assignment of experimental families to the 11 different *NIT* programs was not made randomly, but instead assignment was made on the basis of "normal income level." Those with low normal incomes tended to be concentrated in the programs with small breakevens. Simple predictions of the experimental sample's behavior from the control sample would not take into account this nonrandom assignment. Since control families were also assigned to normal income levels, predictions of participation patterns were made for control families within each normal income level. The final prediction was a weighted average of the predicted probabilities over the eight normal income levels, the weights reflecting the distribution of experimental families in that particular *NIT* program. Mathematically, let $p(i,j,k)$ = the probability that a control family with normal income level i follows the kth participation pattern in the jth *NIT* program, and let $w(i,j)$ = the percentage of experimental families in the jth *NIT* program who have normal income level i. Then the percentage of experimental families predicted to follow the kth pattern in the jth program is

$$\sum_{i=1}^{8} p(i,j,k)w(i,j).$$

[12]The statistic computed is distributed as χ^2_7, approximately. Sidney Siegel (1956) presents conditions under which this approximation is close.

TABLE 1—ACTUAL (PREDICTED) PARTICIPATION PATTERNS OF WHITE FAMILIES
(Percent of Sample)

Break-Even Income	Years in which Families Participated: None	1	2	1,2	3	1,3	2,3	1,2,3	Number of Observations	χ^2_7
\$ 5,430	66.7	13.3	0.0	6.7	3.3	3.3	3.3	3.3	30	4.17
	(58.7)	(11.5)	(1.3)	(3.6)	(6.1)	(2.1)	(3.6)	(13.2)		
5,801	38.7	6.5	3.2	6.5	9.7	3.2	6.5	25.8	31	6.54
	(43.1)	(20.1)	(2.4)	(4.4)	(5.1)	(2.0)	(2.5)	(20.5)		
6,850	32.4	20.3	4.1	4.1	6.8	2.7	5.4	24.3	74	11.40
	(50.5)	(12.7)	(1.8)	(4.8)	(5.4)	(4.4)	(3.2)	(17.2)		
7,366	21.1	7.9	0.0	18.4	5.3	2.6	7.9	36.8	38	11.45
	(25.2)	(19.4)	(2.3)	(6.7)	(4.2)	(6.6)	(5.7)	(29.8)		
7,600	20.6	17.6	5.9	8.8	5.9	5.9	4.4	30.9	68	8.92
	(31.1)	(16.0)	(1.5)	(6.2)	(3.7)	(7.4)	(4.1)	(28.4)		
8,000[a]	34.4	9.8	4.9	4.9	3.3	1.6	6.6	34.4	61	9.31
	(37.2)	(15.0)	(1.8)	(6.2)	(5.8)	(3.7)	(2.1)	(28.1)		
8,000[b]	25.6	20.9	7.0	11.6	4.7	7.0	0.0	23.3	43	11.44
	(31.0)	(15.9)	(1.5)	(5.5)	(4.9)	(4.9)	(2.3)	(34.0)		
9,600	17.4	8.7	2.9	7.2	4.3	1.4	13.0	44.9	69	21.26
	(22.7)	(14.3)	(3.3)	(6.2)	(2.6)	(6.2)	(2.5)	(40.8)		
10,343	22.4	9.2	3.9	5.3	2.6	1.3	5.3	50.0	76	7.09
	(17.2)	(11.3)	(2.1)	(9.0)	(3.8)	(5.4)	(2.6)	(48.5)		
11,200	12.5	6.3	0.0	6.3	6.3	0.0	3.1	65.6	32	4.13
	(12.1)	(8.6)	(2.3)	(8.8)	(2.7)	(4.4)	(2.1)	(59.0)		
12,000	11.6	10.1	1.4	15.9	7.2	2.9	2.9	47.8	69	18.81
	(7.8)	(5.7)	(0.3)	(7.8)	(2.0)	(3.5)	(2.4)	(70.5)		

Note: Predictions are shown in parentheses.
[a]This program does not have a declining tax rate ($G = 5600$, $t = 0.7$).
[b]This program does have a declining tax rate ($G = 4800$, $t = 0.8$, rate of decline $= 0.000025$).

of \$5,430, one cannot reject the hypothesis that the actual and predicted distribution of families among participation patterns is the same, since the χ^2 statistic does not exceed the critical value at the 95 percent confidence level.

Examination of Table 1 indicates that this simple model of participation behavior does a good job of explaining the data. Statistically, the hypothesis that the experimentals are behaving as predicted by the controls can be rejected in only two of the eleven programs. In the programs with break-even income levels of \$9,600 and \$12,000, the χ^2 statistic is greater than the critical value of 14.07. However, there is reason to doubt the adequacy of this simple model. First, in the intermediate range of break-even levels, there tends to be underprediction of participation. For example, for the pattern of participating, in all three years, this underprediction ranges from 1.5 percent to 7.1 percent. Second, if we aggregate the 11 χ^2 statistics, the sum is 114.5 which is large enough to cast doubt on the simple model for the entire data set.[13] This simple model, which relies only on the break-even level for prediction of behavior is nested within the model which includes economic behavior. The test I am using to judge this simple model is not powerful against certain alternatives. For example, if any economic effect were concentrated in a subset of 11 *NIT* programs (perhaps those with the highest implicit tax rates), we would expect rejection of the null hypothesis only for a few break-even levels. Acceptance of this simple statistical model at this point would lead us to ignore any potential economic behavior. Thus I proceed to test the more complicated model of participation behavior as expressed in equation (3).

[13]If we view the χ^2 statistics as being independent, the sum is χ^2_{77}. The independence assumption is invalid since the entire control sample is used to make all the predictions about experimental behavior. Nevertheless this number gives us an indication that overall the model is not behaving particularly well.

TABLE 2

Sample	Number of Observations	Minimum χ^2 Estimates		Jackknifed Estimates	
		e	F/G	e	F/G
Whites	958	0.134	0.021	0.131 (0.016)	0.017 (0.019)
Blacks	646	−0.080	0.073	−0.085 (0.057)	0.056 (0.047)
Chicanos	396	−0.223	−0.174	−0.224 (0.009)	−0.163 (0.058)
Whites in Seattle	445	−0.372	−0.156	−0.323 (0.668)	−0.143 (0.132)
Whites in Denver	513	0.078	0.090	0.090 (0.112)	0.093 (0.027)
Three-Year Families	776	0.115	0.016	0.047 (0.438)	0.030 (0.058)
Five-Year Families	549	−0.019	0.047	0.058 (0.237)	0.016 (0.112)

The model summarized in equation (3) depends on two unknown parameters e and F/G, and any predictions of experimental families behavior require knowledge of the underlying income distribution, $f(\cdot)$. If $f(\cdot)$ were known, e and F/G could be estimated using a probit-like maximum likelihood scheme. Such estimating procedures are known to be very sensitive to distributional assumptions and there is a simpler approach. In Table 1, I compared the actual participation patterns of the experimental families to the predicted patterns of the controls. These predicted patterns were generated assuming that the rule for participation was a function of only the break-even level. In particular, a control family was said to be a predicted participant in period s if $\ln(Y_s) < \ln(G/t)$; implicitly it was assumed $e = 0$ and $F/G = 0$. I could have calculated predictions using equation (3) and any values of e and F/G. Alternatively, I could choose as estimates of e and F/G the values that provide the closest fit between predicted and actual participation patterns. I do so, using the sum of χ^2 statistics over all 11 break-even levels as the criterion function that measures closeness of fit.[14]

The results of the estimation are presented in Table 2.[15] The first column lists the subset of the sample for which the estimation was done. The corresponding number of families is listed in the second column. The third column shows the result of the minimum χ^2 estimates. For example, the white families have an estimated compensated elasticity of labor supply of 0.134 and a very weak state dependence effect of only 0.021. Whites show the largest sensitivity to marginal tax rates and blacks show the most pronounced state dependence effect. The Chicanos show a negative state dependence effect indicating they move in and out of subsidization with ease. The families in Denver show the largest behavioral reactions to the program. Families on the long-exposure program (five years) have a smaller value of the elasticity and a larger state dependence effect. The point estimates presented in Table 2 describe the behavior one would conjecture on the basis of Table 1 alone: a small sensitivity to tax rates indicating an underprediction of participation and a small welfare trap leading to more persistence than would otherwise be

[14]I am grateful to Hal White for initially suggesting this general approach. Several other criteria functions were tried with no substantive qualitative or quantitative differences. See my dissertation.

[15]The estimates were calculated using the simplex search method and a method due to M. J. D. Powell (1964) that does not rely on first derivatives of the criterion function. The simplex method is described by J. A. Nelder and R. Mead (1965). The actual routines used are contained in the optimization package GQOPT.

predicted. The conclusion is that the important reason for persistent participation is correlation in earnings.

The negative values of the estimates of e and F/G in Table 2 are disturbing at first glance because they do not square with economic theory. However, if the true values of e and F/G were close to zero and there were some noise in the estimation procedure, one would expect to see some negative estimated parameter values depending on the sample on which the estimate was based. Repeated estimation, based on different random samples would yield estimates which on average equalled the true value of the parameters.[16] A weighted average of e and F/G over the three ethnic categories yields average estimates of -0.005 and -6.0×10^{-4}, respectively. Another way to gauge the precision of these minimum χ^2 estimates is to compute standard errors using a jackknife procedure described by Frederick Mosteller and John Tukey (1977). The experimental sample was broken into 20 groups and the pseudovalues were computed by deleting each twentieth of the sample in turn. The estimated variable is the sample variance of those pseudovalues.[17] This procedure also yields jackknifed estimates of the parameters of interest. For whites, these estimates are $e_j = 0.131$ and $(F/G)_j = 0.017$ where the subscripts denote jackknifing. The corresponding standard errors are $s_e = 0.016$ and $s_{F/G} = 0.019$. The jackknifed estimates for the other subpopulations indicate that the actual point estimates in Table 2 are sensitive to outlying observations and the precision of estimation is not great. With these precision estimates in hand we see that, with the exception of the Chicano population, the negative estimates are insignificant, and for the most part the behavioral estimates are insignificant altogether.

The minimum χ^2 estimates of the key parameters indicate that earnings correlation is the culprit in the persistent participation of families in these *NIT* programs. There is some evidence for a small positive compensated elasticity of labor supply, but the welfare trap does not seem to be an empirically compelling explanation for persistent participation. The numerical results obtained do not seem very precise. Since the estimation method I used did not rely on any distributional assumptions, the tests I have performed are not very powerful. I can attempt to remedy this problem by making some specific distributional assumptions that allow me to perform some standard and hopefully more powerful statistical tests of the model.

The maximum-likelihood approach to estimation is a straightforward extension of the theory I have developed. Using the criterion function in (3) and assuming some multivariate distribution of income $f(\cdot)$, the likelihood function for the families in the experimental sample is simply the product of the appropriate multivariate integral of $f(\cdot)$. If $f(\cdot)$ is log normal, this is a multivariate probit estimation. The likelihood function for the experimental sample can be augmented by the corresponding likelihood function for the control sample which is simply the product of the density function $f(\cdot)$ evaluated at actual earnings.[18]

[16]The problems with theoretical inequality boundaries on empirical estimates have been discussed by Michael Lovell and Edward Prescott (1970).

[17]For any parameter θ_i, the parameter estimate is the average of five pseudovalues computed from the optimization routines:

$$\theta_i = \frac{1}{5}\sum_{j=1}^{5}\theta_{ij}^*$$

where θ_{ij}^* is the jth pseudovalue of the ith coefficient. The standard error s_* is computed as $s_*^2 = s^2/5$, where

$$s^2 = \left[\Sigma\theta_{ij}^{*2} - \tfrac{1}{5}(\Sigma\beta_{ij})^2\right]/4.$$

See Mosteller and Tukey.

[18]More precisely, given the density function $f(\cdot|\mu,\sigma)$ where μ is a time variant vector that depends on family characteristics such as race and location; i.e., $\mu = X\beta$, which is a typical earnings function. The likelihood function for the N_e experimental families is simply the product of the appropriate multivariate integral given each family's characteristics and participation pattern:

$$\mathscr{L}E = \prod_{j=1}^{N_e} P_J,$$

where P_J is the appropriate integral of $f(\ )$. The likelihood function of the control subsample is simply the product of the density function evaluated at actual

TABLE 3—LABOR SUPPLY AND STATE DEPENDENCE RESULTS FOR VARIOUS DEMOGRAPHIC GROUPS[a]

Demographic Group	Number of Observations	Compensated Elasticity of Labor Supply (e)	State Dependence Parameter (F/G)
Whites	958	0.968	0.352
	(4.56)	(5.50)	
Blacks	646	0.284	0.432
	(1.22)	(5.86)	
Chicanos	396	0.429	0.080
	(1.27)	(0.82)	
White Families:			
Denver Residents	513	1.367	0.561
		(4.15)	(5.47)
Seattle Residents	445	0.434	0.069
		(1.75)	(0.97)
Three-Year Controls	776	0.900	0.392
		(3.62)	(5.25)
Five-Year Controls	549	1.219	0.343
		(3.71)	(3.14)

[a]Asymptotic t-statistics are shown in parentheses.

Table 3 presents the important maximum-likelihood parameter estimates for various subsets of the sample. I have assumed a log-normal distribution of earnings and the specification assumes a mean which is conditional on the period of observation, pre-experimental earnings, and the normal income level used for assignment to the various *NIT* programs. The covariance matrix is unconstrained. (In my dissertation, a detailed explanation of the estimation procedure is given along with a full set of parameter estimates.) The estimates in Table 3 are startling in light of those calculated earlier. For example, the first line of Table 3 shows that, for white families, the maximum-likelihood compensated elasticity is estimated to be 0.968, as opposed to 0.131 using minimum χ^2 squared deviations. The state dependence effect increases for white families from 0.017 to 0.352. These estimates would lead us to the conclusion that these families are showing a strong reaction to the marginal tax rates and there is a pronounced welfare trap, which is contrary to the conclusions drawn from the minimum χ^2 estimates. Which set of estimates are we to believe?

The maximum-likelihood estimates depend on a specific functional form of the density function. The assumption of log-normality of earnings is arbitrary. Using methods devised by Halbert White (1979), the normality of log earnings in this sample was tested accounting for the truncation of the sample that was done prior to the experiment.[19] Using two different methods, log earnings failed the normality test at the 90 percent confidence level. White indicates that if the test is at all biased, it is in favor of the normality assumption. Other estimates were computed using alternative distributional assumptions. Using the logistic function, the compensated elasticity was found to be 0.485 for white families and significantly different from zero; the state dependence parameter was 0.155. The maximum-likelihood estimates from a multivariate discrete choice model are very sensitive to distributional assumptions. In this instance, specifying a particular functional form has radically altered

earnings:

$$\mathscr{L}C=\prod_{j=1}^{N_c} f(y|X_j\beta,\sigma)$$

where N_c denotes the number of control families. The likelihood for the entire observed sample is then $\mathscr{L}_E \times \mathscr{L}_c$.

[19]See White, and White and Lawrence Olson (1981).

TABLE 4—ACTUAL (PREDICTED) PARTICIPATION PATTERNS OF WHITE FAMILIES USING MAXIMUM-LIKELIHOOD PARAMETER ESTIMATES: $e = 0.968$; $F/G = 0.352$
(Percent of Sample)

Break-Even Income	Years in which Families Participated								Number of Observations	χ^2_7
	None	1	2	1,2	3	1,3	2,3	1,2,3		
$ 5,430	66.7	13.3	0.0	6.7	3.3	3.3	3.3	3.3	30	3.63
	(58.7)	(7.2)	(0.6)	(4.3)	(6.1)	(0.6)	(4.2)	(18.2)		
5,801	38.7	6.5	3.2	6.5	9.7	3.2	6.5	25.8	31	16.31
	(45.7)	(10.0)	(2.0)	(2.2)	(5.1)	(0.0)	(4.5)	(30.6)		
6,850	32.4	20.3	4.1	4.1	6.8	2.7	5.4	24.3	74	44.58
	(52.7)	(3.0)	(0.8)	(2.6)	(5.6)	(0.8)	(4.2)	(30.3)		
7,366	21.1	7.9	0.0	18.4	5.3	2.6	7.9	36.8	38	8.13
	(29.7)	(8.3)	(2.3)	(8.6)	(6.2)	(0.4)	(6.2)	(38.3)		
7,608	20.6	17.6	5.9	8.8	5.9	5.9	4.4	30.9	68	39.57
	(38.9)	(4.0)	(1.4)	(5.4)	(5.5)	(0.8)	(4.8)	(37.7)		
8,000[a]	34.4	9.8	4.9	4.9	3.3	1.6	6.6	34.4	61	20.39
	(39.0)	(3.1)	(0.7)	(2.4)	(5.8)	(0.4)	(2.8)	(45.9)		
8,000[b]	25.6	20.9	7.0	11.6	4.7	7.0	0.0	23.3	43	73.38
	(38.4)	(4.1)	(0.2)	(2.2)	(4.6)	(0.4)	(5.4)	(44.8)		
9,600	17.4	8.7	2.9	7.2	4.3	1.4	13.0	44.9	69	33.89
	(34.8)	(3.3)	(0.0)	(2.4)	(3.8)	(0.8)	(4.0)	(49.5)		
10,343	22.4	9.2	3.9	5.3	2.6	1.3	5.3	50.0	76	32.32
	(26.6)	(1.8)	(0.0)	(1.8)	(3.8)	(0.2)	(7.0)	(58.7)		
11,200	12.5	6.3	0.0	6.3	6.3	0.0	3.1	65.6	32	10.83
	(21.5)	(0.9)	(0.3)	(2.1)	(2.7)	(0.0)	(5.0)	(67.5)		
12,000	11.6	10.1	1.4	15.9	7.2	2.9	2.9	47.8	69	63.59
	(15.2)	(1.6)	(0.6)	(1.9)	(2.4)	(0.0)	(4.5)	(73.7)		

Notes and footnotes: See Table 1.

the conclusion. Because I have experimental data, I can check the distributional assumption which is a luxury not given to many economic investigators. The White tests lead us to reject normality, but do not lead to any reasonable alternative specification. The search for the correct maximum-likelihood estimates becomes a very expensive guessing game.

The maximum-likelihood estimates are potentially "better" estimates than the minimum χ^2 estimates if they do a better job of explaining the observed data. Because we have experimental data available, we can use the control group to gauge the explanatory power of the *ML* estimates. In Table 4, I recomputed predicted participation patterns for the white control families using the *ML* point estimates. Contrast these predicted patterns (shown in parentheses) with the actual patterns directly above, and it is easy to see that the *ML* estimates do not provide a cogent explanation of the behavior we observe in the data. For example, in the $9,600 break-even program, the *ML* model leads to substantial overprediction for the extreme patterns of always participating or always not participating, and underprediction for the mixed patterns. Because the *ML* estimates are generated using a seemingly incorrect distributional assumption, the simpler minimum χ^2 estimates do a much better job of explaining the observed data. The latter estimates do not allow us to perform tests as powerful as those that could be made if we knew the correct density function. To compute those estimates is a difficult if not intractable problem.

III. Conclusion

In attempting to explain participation in welfare programs, this study has made several observations, both empirical and methodological, which I now review.

First, it was noted that participation in *NIT* programs is determined by the pattern of earnings in the population. Persistent par-

ticipation will be observed among families if earnings are highly correlated. The model accounted for this reason for persistent participation, but also incorporated potential economic reaction to the subsidy program. Because of the implicit wage tax, there was the potential for substitution into leisure and, moreover, there might be a state dependence effect that leads to increased participation over time. The idea of state dependence has been discussed widely in the econometric and labor economics literature. In this instance, the existence of state dependence is an important question for policy purposes. If there is state dependence, that is, if a welfare trap does exist, it means that economic incentives to work are steadily declining because of welfare programs. I would conclude that the government should limit the length of families' eligibility so they do not become entrapped into dependence on the government's generosity. If there were no welfare trap and we are concerned with persistence in participation, we must look to other more structural remedies for the problem.

The empirical finding of this study is that persistence is explained by correlation in earnings and the evidence pointing towards a welfare trap is at best weak. There is some indication that there is tax rate sensitivity, but the compensated elasticities are imprecisely estimated. The magnitudes of the elasticities are similar to those found elsewhere.[20] This study supports the notion that policymakers should worry about the social structures that entrap the poor into persistent low earnings rather than portraying short-term disincentives to work inherent in the welfare system as the culprit.

Finally, for the empirical labor economist, the study has made a methodological observation. The use of discrete choice maximum-likelihood procedures in estimating the parameters of our model would have led us astray if the specification of the density function was incorrect. A simpler method that did not depend on any distribution was used to find estimates that explained the data in a much more cogent way. Applied econometricians need to consider carefully the methods used, the implicit assumptions contained in the implementation of those methods, and the ability of the resultant model to describe the observed data.

This study has attempted to explain the patterns of welfare participation. Two avenues for future research should be clear. First, a full intertemporal model of labor force behavior should be incorporated into the model of discrete behavior. Second, the econometric methodology must be studied more carefully if future research will be expected to handle problems like this in a careful and correct way.

[20] See Ashenfelter (1983).

REFERENCES

Ashenfelter, Orley, "Discrete Choice in Labor Supply: The Determinants of Participation in the Seattle-Denver Income Maintenance Experiments," *Journal of the American Statistical Association*, September 1983, *78*, 517–25.

______, "The Labor Supply Response of Wage Earners," in John Palmer and J. A. Pechman, eds., *Welfare in Rural Areas: The North Carolina-Iowa Income Maintenance Experiment*, Washington: The Brookings Institution, 1978, 109–38.

Harrington, Michael, *The Other America: Poverty in the United States*, New York: Macmillan, 1962.

Keeley, Michael C. and Robins, Philip D., "Experimental Design, the Conlisk-Watts Assignment Model, and the Proper Estimation of Behavioral Response," *Journal of Human Resources*, Fall 1980, *15*, 480–98.

Lovell, M. C. and Prescott, E., "Multiple Regression with Inequality Constraints: Pretesting Bias, Hypothesis Testing and Efficiency," *Journal of the American Statistical Association*, June 1970, *65*, 913–25.

Mosteller, Frank and Tukey, John W., *Data Analysis and Regression*, Reading: Addison-Wesley, 1977.

Nelder, J. A. and Mead, R., "A Simplex Method for Function Minimization," *Computer Journal*, January 1965, *7*, 308–13.

Plant, Mark W., "A Note on the Trivariate Probit Model as a One-Factor Scheme,"

unpublished paper, Princeton University, September 1980.

_____, "Participation Over Time in the Seattle-Denver Maintenance Programs," unpublished doctoral dissertation, Princeton University, January 1982.

Powell, M. J. D., "An Efficient Method for Finding the Minimum of a Function of Several Variables Without Calculating Derivatives," *Computer Journal*, July 1964, *7*, 155–62.

Rea, Samuel A., "Investment in Human Capital Under a Negative Income Tax," *Canadian Journal of Economics*, November 1977, *10*, 607–20.

Robins, Philip K. and West, Richard W., (1980a) "Labor Supply Response Over Time," *Journal of Human Resources*, Fall 1980, *15*, 524–44.

_____ **and** _____, (1980b) "Program Participation and Labor Supply Response," *Journal of Human Resources*, Fall 1980, *15*, 499–523.

Siegel, Sidney, *Nonparametric Statistics*, New York: McGraw-Hill, 1956.

White, Halbert, "Specification Tests for the Tobit Model," Discussion Paper No. 79–22, University of California, San Diego, 1979.

_____ **and Olson, Lawrence,** "Conditional Distributions of Earnings, Wages and Hours for Blacks and Whites," unpublished paper, University of California, San Diego, 1981.

Race and Human Capital

By JAMES P. SMITH*

While human capital has been used with some success to analyze recent changes in racial income differences, scholars have repeatedly pointed to a major empirical problem that appears to severely limit the historical relevance and scope of skill-based theories as applied to racial questions. The challenge they raise is legitimate. Put simply, if measured skill disparities between the races narrowed throughout the twentieth century, why did income ratios first begin to converge in the 1960's?

In this paper, I address this question relying on some unexploited census data by race on education, literacy, occupations, and income. Using these data that begin with the 1890 Census, I present new estimates of age-specific relative income positions of black men for all postslavery birth cohorts. In addition to reconciling the apparently inconsistent skill and income series, these income ratios offer a very different historical record than many economists believe to have been the case. To cite a prominent example, Gunnar Myrdal's classic work (1944) saw the economic position of his contemporary black America not only as dismal, but made even more so by its sense of hopelessness, given the absence of any hint of progress or change. While Myrdal's pessimism is understandable, it appears that even in his day seeds had long been sown that were already permanently altering and improving the relative economic status of black men.

I. Background

The success of human capital explanations for racial income differences has historically been uneven at best. Morton Zeman's 1955 dissertation, a detailed and careful analysis of the 1940 Census, served as the standard empirical reference until the micro-level 1960 Census tapes became available. Zeman evaluated the importance of human capital's two main faces: years of schooling and on-the-job training. While incomes of men of both races increased with schooling and age, Zeman discovered that the rate of increase across either dimension was considerably higher for white men. Whether the added skill involved an extra year in school or another on the job, black-white income ratios declined by one-third of 1 percent for every 1 percent rise in white income. Racial income differences within schooling or age groups dominated racial differences in the amount of human capital possessed. The elimination of racial differences in schooling or age would have reduced the southern racial income gap by less than one-quarter and that in the North by about 5 percent. Essentially, Zeman's work implied that giving blacks the same amount of schooling as whites would do little to alter racial income disparities. The problem, rather, was different pay for the same skill and differential racial payoffs to acquiring more skill.

Throughout the 1950's, events appeared to confirm Zeman's research. Studies based on the 1950 and 1960 Censuses indicated that, if anything, there may have been a deterioration in the relative economic position of blacks (Alan Batchelder, 1964). The initial micro-level studies based on the 1960 Census by Giora Hanoch (1965), Lester Thurow (1969), and others did little to amend this view. Their work painted a consistent picture of low returns to schooling for blacks as well as a sharp deterioration in relative black economic potential over job careers. In light of these findings, the emphasis in the economic literature concentrated on reasons for racial differences in the value of skill.[1]

*Senior Economist, The Rand Corporation, 1700 Main Street, Santa Monica, CA 90406. Comments by Gary Becker, Robert Hauser, Jacob Mincer, and Ted Schultz are gratefully acknowledged. I am indebted to the oftentimes heroic assistance of Chris Witsberger. The research received support from the Ford Foundation and the National Science Foundation.

[1]On the schooling dimension, market discrimination against black skilled labor and government discrimination resulting in lower quality of black schools were the principal competing factors advanced to explain de-

TABLE 1—YEARS OF SCHOOLING OF U.S. MALES BY YEARS OF BIRTH

Birth Years	Age in 1973	Black	White	Difference
1907–16	57–66	7.1	10.6	3.5
1917–26	47–56	8.6	11.6	3.0
1927–36	37–46	10.1	12.2	2.1
1937–46	27–36	11.4	12.9	1.5
1947–51	22–26	11.9	13.0	1.1

Source: Robert Hauser and David Featherman (1978).

Two developments led to more emphasis being assigned to human capital factors during the late 1960's and early 1970's. First, there was a sustained and impressive rise in black-white income ratios. Second, micro analysis based on the 1967 Survey of Economic Opportunity and 1970 Census indicated that rates of return to education, especially at upper schooling levels, appeared as high for blacks as for whites (Finis Welch, 1974). In addition, reanalysis of the 1960 Census revealed that first-generation studies had significantly understated rates of return to education for blacks.[2] On the issue of job-related investments, Welch pointed out that cross-sectional data showing declining black-white income ratios with age may not reflect life cycle factors at all if changes across birth cohorts are larger for blacks than whites. In joint work, Welch and I (1977) provided evidence that over recent years, black male incomes did not fall behind those of comparable whites as their careers evolved, lending some support to a cohort interpretation for at least recent cross sections.

The historical validity of an emphasis on human capital factors has rightly been questioned (see Orley Ashenfelter, 1977; Henry Levin, 1978; and William Darity, 1982).[3]

TABLE 2—RATIOS OF MEAN PERSONAL INCOME OF BLACK AND WHITE MALES, 1948–78

Year	Ratios	Year	Ratios
1948	.530	1968	.612
1953	.533	1973	.643
1958	.538	1978	.669
1963	.566		

Source: U.S. Bureau of the Census, *Consumer Income*, Series, P-60.

The basic issue raised is illustrated in Table 1 which summarizes trends in educational attainment by race and post-World War II changes in racial income differences. The twentieth-century schooling story is clear. With each successive birth cohort, racial differences in education levels narrowed, until a 3.5-year gap for the oldest cohort was reduced to only a 1.1-year advantage for whites in the youngest cohort. As new cohorts entered the labor force over the last forty years, the competitive human capital disadvantage

clining black-white income ratios with schooling. On the job investment side, discrimination-based theories dominated, either using the language of denying blacks access to jobs with human capital growth or simply confining blacks to secondary labor markets.

[2]Apparently, due to the use of group data and an age control rather than experience. (See my article with Welch, 1977.)

[3]While the basic point remained the same, the language used became progressively stronger and the problem was viewed as increasingly lethal. "Where does this leave us? In my view, it leaves us with a very considerable puzzle. Attributing the steady increase in relative earnings to the gradual increase in their skills (as measured, say, by schooling) requires an explanation of why the gradual increase in skills of black men had so little effect on their relative earnings before the mid 1960's" (Ashenfelter, p. 297). "Thus, the human capital interpretation seems to be applied in an *ad hoc* fashion to 1964 and beyond while ignoring the lack of earnings equalization during the first half of this century when the trend towards educational equalization was much more dramatic" (Levin, p. 28). "Human capital differences, therefore, are offered only as the dominant explanation for racial income differences for a limited episode—the 1960s and later. Usually, the theory is not claimed by its proponents to be historically general in applicability. ...The human capital approach may well be a dead letter" (Darity, p. 80).

TABLE 3—MEAN SCHOOLING LEVELS BY BIRTH COHORT (YEARS OF SCHOOLING)

Birth Cohort	White Males	Black Males	White Females	Black Females
1951–54	12.64	11.82	12.70	12.24
1946–50	12.68	11.93	12.45	11.86
1941–45	12.32	11.25	12.14	11.33
1936–40	12.00	10.46	11.81	10.89
1931–35	11.69	9.78	11.52	10.37
1926–30	11.38	9.11	11.33	9.87
1921–25	11.14	8.44	11.12	9.03
1916–20	10.74	7.65	10.79	8.36
1911–15	10.15	6.75	10.36	7.70
1906–10	9.72	6.26	10.02	7.16
1901–05	9.19	5.72	9.45	6.46
1896–1900	8.74	5.42	8.96	6.03
1891–95	8.18	4.96	8.42	5.52
1886–90	7.74	4.72	8.11	5.13
1881–85	7.56	4.38	7.95	4.67
1876–80	7.44	4.11	7.88	4.27
1871–65	7.22	3.56	7.58	3.59
1866–70	7.07	3.06	7.45	2.89
pre–1865	6.76	2.37	7.13	1.99

of blacks appears to have continuously diminished.

This was not so for income. Before 1960, there was little change in relative income by race. Black-white male income ratios did rise sharply after the mid-1960's and have continued to do so for the last fifteen years.[4] See Table 2.

The appreciable rise in relative black income that appears to begin suddenly during the mid-1960's is roughly coincident with the passage of the 1965 Civil Rights Act. This correspondence in timing is the key piece of evidence for those who assign an almost exclusive role to antidiscrimination legislation (Richard Freeman, 1973). Add to this the basically negative results for human capital factors in studies based on the 1940 and 1950 Censuses and the issue is joined: side by side with steady long-term convergence of human capital characteristics, why is there silence on the income front?

II. Schools

To address this question, I begin with a summary of my historical education data.[5] Table 3 presents new estimates of average schooling completed by five-year birth cohorts,[6] starting with persons born in the first

[4]Indeed, a common initial reaction to the 1960's acceleration was to view it as an historical aberration, a short-run by-product of the boom conditions of that decade. As the improvement continued unabated into the late 1970's, when business cycle conditions were less than robust, the idea that this was only a transitory improvement was put aside.

[5]Between 1870 and 1930, the census included questions on illiteracy, defined as the ability to read and write English or another language. Beginning with the 1940 Census, the literacy questions were replaced with questions on educational attainment.

[6]My estimates are based on published census data from the 1940 and 1970 Censuses and the 1979 *Current Population Survey* (*CPS*). Published data include single year of schooling distributions by five-year age groups separately by race and sex. Each cross-sectional census provides an estimate of schooling distributions for a set of birth cohorts. By linking and blending these five cross sections, one can span the birth cohorts listed in Table 3. Figures on education by birth cohort are averages of distributions across census years, so that data for a single birth cohort could potentially come from five cross sections. This averaging did not significantly alter the time-series description that one would obtain within any census year. As is well known, birth cohort tracking across census years shows unbelievably large increases (beyond any reasonable age of school attendance) in average schooling levels. This problem is particularly acute for blacks. This education inflation is often attributed to exaggeration of schooling accomplishments as education norms in society rise.

TABLE 4—RACIAL DIFFERENCES IN MEAN SCHOOLING LEVELS (ADDITIONAL YEARS OF SCHOOLING OF WHITES)

Birth Cohort	Men	Women
1951–54	.83	.46
1946–50	.75	.60
1941–45	1.07	.81
1936–40	1.54	.92
1931–35	1.91	1.15
1926–30	2.27	1.47
1921–25	2.71	2.09
1916–20	3.09	2.44
1911–15	3.41	2.69
1906–10	3.46	2.86
1901–05	3.46	3.00
1896–1900	3.32	2.93
1891–95	3.23	2.90
1886–90	3.02	2.98
1881–85	3.18	3.28
1876–80	3.33	3.62
1871–75	3.67	4.00
1865–70	4.01	4.55
pre–1865	4.39	5.13

five years after the Civil War. To isolate differential secular trends, Table 4 lists racial differences in school completion for each birth cohort. Finally, a more complete characterization of schooling distributions is given in Table 5, evaluated for selected birth cohorts that represent key historical turning points.

The 120-year post-Civil War period turns out not to be a simple story of steady and constant narrowing of human capital differences between the races, as backward extrapolation of twentieth-century data would have led one to expect. For those born in this century, as we have seen earlier, the underlying trend was indeed one of convergence. Similarly, racial differences in mean schooling levels diminished sharply for both sexes during the first quarter-century after the Civil War. Matters took a different course, however, for the generations born between 1886 and 1905. During this period, racial disparities in school completion expanded for males and stabilized for women. This 20-year reversal has had important implications for the economic history of blacks right to the present. And as these generations slowly pass through the 10-year records of the census, they affect our interpretation of income trends.

A convenient way of summarizing the historical data is to separate this 120-year span into three distinct segments: the initial post-Civil War phase of convergence (1865–86), the period of divergence (1886–1905), and finally, the more familiar narrowing of schooling differences during the twentieth century.

A. *1865–1886*

The story of black education in the South before the Civil War is simple indeed. Slave insurrections in the early part of the nineteenth century led to the legal proscription of formal instruction. By 1835, every southern state had a law prohibiting the schooling of slaves, and some even forbade instruction of freedmen. At the start of the Civil War, consequently, 95 percent of black southerners were completely illiterate. At its end, the differences between the races in school completion that I describe in Table 3 were enormous, with whites of either sex finishing 4 to 5 more years of school than blacks.

Although schooling levels would remain low for blacks throughout the nineteenth century, the first 20 years after the Civil War saw the effective origin of mass Negro education in the South. Between 1870 and 1880, the number of black southern children enrolled in school rose by almost 500 percent as the proportion of young blacks in school increased from 9 percent to almost one-third (Welch). In contrast, during this 20-year subperiod, changes in schooling distributions across white male birth cohorts listed in Table 5 were relatively modest, especially compared with what was to come. The modal level of school completion was *exactly* 8 years of grade school. Fully one-third of all white men born in this period received their grade school diploma but never advanced beyond that point. Even by the 1886–90 birth cohorts, only 30 percent of white males went beyond elementary school, and less than one-fifth earned a high school diploma.

Throughout these first two decades after the Civil War, the effective margin for black children was whether to attend school at all, and if so, whether to proceed as far as the fourth or fifth grade. However, we shall see

TABLE 5—SCHOOLING DISTRIBUTIONS OF SELECTED MALE BIRTH COHORTS

	Years of Schooling						
Birth Cohorts	0	< 5	< 9	8	< 12	12	12+
A. Black							
Pre–1865	45.9	81.6	96.5	4.2	98.6	0.9	1.4
1866–70	33.8	73.5	94.9	6.1	97.1	1.3	1.6
1886–90	15.7	54.1	88.0	10.1	93.2	3.5	3.3
1906–10	7.1	35.1	75.9	12.6	88.1	7.3	4.6
1926–30	2.3	11.7	43.7	12.4	67.6	20.2	12.1
1946–50	1.0	1.7	8.4	3.9	26.3	46.2	27.5
B. White							
Pre–1865	7.7	25.3	83.9	34.8	88.9	5.3	5.8
1866–70	6.6	22.4	81.2	35.6	87.4	6.1	6.5
1886–90	5.6	18.9	70.6	29.7	81.5	9.0	9.8
1906–10	1.3	6.1	44.7	23.2	64.1	19.0	16.1
1926–30	0.8	3.0	20.7	10.3	41.0	32.3	26.7
1946–50	0.5	1.9	7.7	2.9	18.8	36.8	44.4

TABLE 6

	White Men		Black Men	
Birth Cohort	Go On	Complete High School	Go On	Complete High School
A. School Continuation Probabilities for Elementary School Graduates				
1916–20	83.1	59.7	75.6	41.2
1906–10	70.4	44.7	65.7	32.4
1896–1900	58.6	36.2	57.0	29.2
1886–90	49.8	31.3	54.3	30.8
1876–80	44.0	29.0	51.4	29.1
B. Increase in Average Schooling by Year of Birth[a]				
	North	South	North	South
1888–1908	2.30	1.75	2.09	2.01
1908–28	1.64	1.58	1.78	1.94

[a]*Source:* Public Use Tapes of 1960 U.S. Census.

that these first 4 years of school were critical in shaping the economic history of blacks, for they would eventually eradicate black illiteracy.

B. *1886–1906*

The next 20 years, 1886–1906, represent a temporary stall during which human capital of blacks, as measured by years in school and quality of schooling, fell behind that of whites. This subperiod roughly coincides with the erosion of black political power in the South as the forces of disfranchisement worked their way through the system. But events in the North were of equal import in causing this reversal.

In the North, the principal development was the spread of public high school education on a wide scale. No longer did the eighth grade constitute the normal end of a school career for a white child. The speed at which white high school attendance and completion became common during this subperiod is illustrated in Table 6, which lists school continuation propensities, conditioned on elementary school graduation.[7]

[7]I have calculated elementary graduate progression rates using 8 years of school. However, the appropriate year to use is ambiguous because historically, elementary schools did not have uniform lengths. Many elementary schools, particularly in the South, contained only 7 grades.

Across the span of years that covers those born in 1886 to those born in 1916, the fraction of white males who continued their education beyond elementary school rose dramatically. By the birth cohort of 1916, more than 8 in 10 white male elementary graduates continued in school, and 60 percent would earn their high school diplomas —almost double the rates that existed 30 years earlier. In contrast, these high school completion rates were stable (at around 30 percent) among blacks between 1886 and 1910.[8] Table 6, panel B, attests to the lead taken by the North in the development of the public high school. Between birth years 1888 and 1908, the largest increases in mean education were achieved by northern-born whites. In contrast, for those born between 1908 and 1928, the largest schooling gains would be registered among southern-born blacks.

Political developments in the South also played a role in causing this period of reversal. With the return to office of conservative Democrats in the elections of 1876–77 and the end of Reconstruction, whatever equivalence had previously existed between the quality of black and white southern schools began eroding, a process that accelerated with disfranchisement.

Starting with Horace Bond (1934), a number of scholars have documented the impact of the loss of black political power on black southern schools. The ultimate legacy was a sharp decline in the relative quality of black southern schools during the late nineteenth century and early part of the twentieth century.[9] There is not space to repeat their documentation here, but I will illustrate with one famous example. In 1886, Mississippi passed a law that allowed counties to divert state funds allocated to black schools. Since state allocations were made to the county independently of pupils' race, the effect of diverting funds to white children on a per capita basis was greatest whenever the ratio of black to white pupils was highest. By 1910, counties with a black majority spent 7 to 30 times as much on white as on black students, while "white" counties spent only 2 to 3 times as much on white pupils. This same Mississippi law also provided for teacher salaries to be related to teacher scores on examinations. From that point (1886), Table 7 demonstrates that black and white teacher salaries drifted apart.

Table 7—Average Monthly Salary of Mississippi Teachers, 1875–95

Year	White	Black
1875	57.5	53.3
1885	28.7	28.6
1890	33.4	23.2
1895	33.0	21.5

Source: Bond.

C. *1906–50*

Because trends in the twentieth century are so widely known, my summary of this subperiod is brief. In terms of schooling quantity, we know that as each new generation of blacks and whites born in this century arrived in the labor force, the difference in schooling achieved by race continuously narrowed. As Tables 3 and 5 show, these changes were never dramatic or sharp. Only for black cohorts born after the depression of the 1930's did high school diplomas become a realistic goal. And, of course, the black college graduate is mostly a product of our own day. The message conveyed by nominal years of schooling is reinforced by data on schooling quality, which tell a clear twentieth-century story of improving relative quality of black schools.[10]

[8]Note that black progression rates were initially higher than those of whites. Those blacks who reached the eighth grade were clearly a very select group. Later in this century, progression rates for high school graduates to college were also higher for black men, a reflection of a similar phenomenon.

[9]Because the date of disfranchisement and the ensuing deterioration in black schools varied across southern states, the southern component of my argument cannot be neatly placed into any single 20-year time frame. However, they roughly coincide with the period in which racial disparities in school quantity expanded.

[10]To cite but two examples, in 1920, black youths attended school only two-thirds as many days per year as white students, but by 1954, there were no real

III. The Effects of Schools

In the remainder of this paper, I will try to document how this history of black schools shaped postslavery swings in the relative economic status of blacks. The first link in the chain of my argument concerns the influence of schools on black literacy.

A. *Literacy*

Table 8 presents illiteracy rates by race for the 100-year period from 1870 through 1969. In each decade, the percentage of blacks recorded by the census as unable to read and write declined significantly, but high rates of illiteracy persisted well into the twentieth century. At the turn of the century, 35 years after the Civil War, almost one-half the black population remained illiterate. Forty years later, at the outbreak of World War II, the fraction was one-tenth. In short, it took 75 years for blacks to equal the basic literacy rate that whites enjoyed at the end of the Civil War.

TABLE 8—PERCENTAGE ILLITERATE BY RACE POPULATION 10 YEARS AND OLDER

Year	All	Whites	Blacks
1870	20.3	11.5	81.4
1880	17.3	9.4	70.0
1890	13.3	7.7	57.1
1900	10.7	6.2	44.5
1910	7.7	5.0	30.4
1920	6.0	4.0	22.9
1930	4.3	2.7	16.3
1940	2.9	2.0	11.5
1952	2.5	1.8	10.2
1959	2.2	1.6	7.5
1969	1.0	0.7	3.6

Source: 1870–1930 from *Fifteenth Census of the United States: 1930 Population*, Vol. 11, *General Report Statistics by Subjects*, ch. 13–"Illiteracy." 1940–69: "Illiteracy in the United States," *Current Population Reports*, Series P-20, Nos. 45 and 217, November 1969.

Table 9 illustrates the effect of schools on literacy and the reason why black illiteracy remained so high well into the twentieth century. In each census year, this table lists by age group the proportion of black men who were illiterate. There is now abundant evidence that literacy is a skill that is almost exclusively acquired in schools. For example, we find that over 80 percent of men with a year of schooling or less were unable to read or write. In contrast, among men who completed the fourth grade, only 7 percent were illiterate.[11] Not surprisingly, then, the patterns in Table 9 indicate a sharply rising proportion of black illiterates as we move from younger to older age groups in any census year. The effects of schools are seen most dramatically in the youngest age groups. Illiteracy rates of these blacks aged 10 to 14 fell rapidly as public schooling spread. The fraction of illiterates was half as high in 1900 as it was in 1880, and it would be cut in half again by 1920.

However, if we follow a birth cohort over their lifetime, Table 9 also demonstrates that the adult black population remained largely untouched by the increased ability of their children to read and write. If a black man reached adulthood without this ability, the chances were over 90 percent that he would never be able to write his name or read a newspaper. The idea that after slavery was abolished large numbers of adult blacks successfully acquired, on their own, elementary reading or writing skills is a myth. Average illiteracy in the black population remained

black-white differences in days attended. Similarly, in 1920, teachers of black students had 1.75 times as many pupils as the average teacher in the country. By 1954, this difference had been reduced to 1.18. See my article with Welch for a more complete summary.

[11] Proportion of males, 35 to 44 years old, illiterate by schooling (1947):

Years of Schooling	Percent
0	87.9
1	82.5
2	52.8
3	27.6
4	7.1

Source: Special Reports of U.S. Census (1948). Because the census asked either years of schooling or literacy, cross classifications for earlier periods are not possible.

TABLE 9—PROPORTION OF BLACK MALES ILLITERATE, BY AGE

Age	1969	1952	1930	1920	1910	1900	1890	1880
10–14			5.3	11.4	18.9	30.1	39.8	66.2
15–19			8.1	14.1	20.3	31.8	42.6	62.7
20–24			12.2	17.0	23.9	35.1	49.3	
25–34	1.3	6.4	13.0	17.9	24.6	39.3	56.8	
35–44		6.6	16.8	23.3	32.3	52.0	70.5	
45–54	5.5	11.5	24.2	34.1	47.0	68.1	80.8	73.1
55–64		18.1	34.4	49.4	63.0	78.4	86.3	
65+	16.7	33.3	55.7	68.3	74.5	85.4	90.2	
15–24	0.6	3.9	10.5	15.5	22.1	33.4	45.6	

Source: See Table 8.

TABLE 10—MALE ILLITERACY RATE BY COHORT

Birth Year	Cohort: Age 10[a]	Cohort: Age 40[b]	Proportion Illiterate: Blacks	Fathers	Whites	Fathers
1840	1850	1880	74.8	> 85.0	8.0	–
1850	1860	1890	60.6	> 85.0	7.7	–
1860	1870	1900	43.0	83.2	6.3	–
1870	1880	1910	27.7	74.8	5.5	8.0
1880	1890	1920	22.0	60.6	5.3	7.7
1890	1900	1930	17.3	43.0	3.4	6.3
1900	1910	1940	14.1	27.7	2.0	5.5
1910	1920	1950	7.5	22.0	1.2	5.3
1920	1930	1960	6.1	17.3	0.7	3.4
1930	1940	1970	3.4	14.1	0.5	2.0
1940	1950	1980	1.7	7.5	0.4	1.2
1950	1960	1990	0.8	6.1	0.3	0.7
1960	1970	2000	0.4	3.4	0.3	0.5

Source: See Table 7.
[a]Year of schooling.
[b]Midway through work career.

high and changed slowly because improvements in the aggregate occurred only as new generations were born, schooled, and eventually replaced the much older, mostly illiterate black population.

The stability of illiteracy rates among the adult population allows us to measure long-term movements across different generations of blacks and whites. This is illustrated in Table 10, in which the year index in the first three columns measures year of birth, year of schooling, and year in which each birth cohort is midway through their labor market career. Assuming a 30-year generational lag, we also match illiteracy rates of fathers with each birth cohort. The first clear signs of the benefits of the black southern schools created during the Reconstruction period appear with the birth cohorts of 1860 and 1870. The fraction illiterate of the 1870 cohort was 27.7 percent, almost 50 percentage points below that of the cohort born 30 years earlier.

For simplicity, if we take a 40-year-old as the average worker in a year, the 1870 cohort would eventually represent the average black male worker in 1910. Fathers' illiteracy, which serves as a crude proxy for family background, may indicate why advances in relative market earnings of blacks were initially so slow. The fathers of our average black workers in 1910 were slaves born in 1840, and three-quarters were illiterate. Thus, the improvements in schools that start with the 1860 birth cohort first appear as im-

TABLE 11—"PSEUDO" LOG EARNINGS FUNCTION OF BLACK MALES AGED 14–64: 1900 CENSUS[a]

Variables	
South Atlantic Resident	−.3729
	(11.2)
South Central Resident	−.4671
	(14.3)
Age Spline:	
14–24	.1437
	(32.9)
25–34	.0156
	(3.79)
35–44	−.0105
	(2.21)
45+	−.0157
	(4.61)
Illiterate	−.1523
	(7.44)
Constant	6.310
	(148.8)
R^2	.465
Sample size	2962

[a] *t*-statistics are shown in parentheses.

provements in family background with the birth cohort of 1890, a group that represents our average black worker in 1930. Fortunately, this long and dark shadow of the past has an optimistic side. Eventually, the narrowing of racial income differences accelerated. Not only did each new generation of blacks have more schooling relative to contemporary whites than their parents did, but also they had parents with more education relative to whites than their parents did.

If schools were able to at least impart basic literacy skills, the next step is to relate literacy to income, a connection that also serves as our transition into the next section. Fortunately, the 1900 micro data file of the U.S. Census enables us to establish the connection.[12] Table 11 contains a pseudo earnings function for black men based on the 1900 Census. The pseudo label is used because earnings were not directly available from the 1900 file. Rather, an earnings amount was assigned to each worker based on his sex, race, age, and occupation. (These earnings weights are described in the next section.) Therefore, by construction, regressors must operate exclusively through occupational assignment. In essence, this income weight gives a metric that summarizes some very complex movements across occupations. In our 1900 sample, 45 percent of all black working men were illiterate. My estimate is that illiteracy depressed their income by 15 percent, a result that is strongly statistically significant. Because it ignores within-occupation income differences associated with literacy, this estimate is certainly an understatement.

B. *Income*

Few would question the central role of black schools in the eventual elimination of black illiteracy. The link between schools and income, however, is far more uncertain and the role of schools in the economic resurgence of black America is more controversial. One difficulty in establishing any historical connection is that reliable income data by race are a relatively recent phenomenon. The first decennial census to include income questions was the 1940 Census. Before that survey, only scattered bits of racial income data are available, based typically on small and nonrepresentative samples. To trace out a truly long-term history, one must of necessity resort to indirect methods of estimating income by race. Fortunately, the census has published occupation distributions of the work force, separately by race and sex beginning in 1890. Most important, these occupation distributions were typically stratified by age.[13]

This age stratification is an unexploited dimension of census occupational data that

[12] The 1900 Census is a nationally representative sample of the population of the United States. The sample consists of 100,438 individuals, representing a sampling rate of 1/760 of the population. This micro sample was assembled by scholars at the Center for Studies in Demography and Ecology, University of Washington, from manuscripts of the 12th Census of the United States.

[13] Since 1890, the only census with no age stratification by race occurred in 1910. In addition, in 1920, only broad age intervals (i.e., 25–44, 45–64) are available. In all other censuses, five- or ten-year age brackets are used. Clearly, the micro files available beginning with the 1960 Census allow maximum flexibility.

permits us to follow different birth cohorts over their labor market careers. Since these data begin in 1890, the first black birth cohort that we can follow from virtually the start to the end of this labor market life was the first generation not born in slavery (1865–70); but even for blacks born in the last 30 years of slavery, we can observe partial slices of their labor force histories.

The task of assembling and linking these occupation data was formidable. The most vexing issue involved the construction of a meaningful and consistent set of occupational groupings across such a long time span. With each successive census, the number of occupational categories varied and the meaning of some occupations, even with the same title, often changed. In particular, three fundamental revisions were made by the census in their occupational coding system—in 1910, 1940, and 1970. Although the census has published historical series by sex, there exist no historical data stratified by race, sex, and age. Therefore, a major component of this research was the construction of an occupational taxonomy that covered the time span 1890 to 1980.

The system I have constructed places black and white men in 133 occupation categories for the entire period from 1890 to 1979.[14] Using this occupational coding system, black-white male income ratios were calculated for each birth cohort at different points in their life cycles. Each occupation was assigned a race, sex, and age-specific average income. Although I have experimented with income weights from different years, Table 12 is based on 1970 Census incomes and the 35–49 age group.[15] Our measure of relative income can change only when the distribution of blacks and whites across occupations changes over time. Within-occupation changes in relative incomes of blacks during the twentieth century will not affect this index, a point to which I return below.

Before discussing the cohort and age-specific detail available in Table 12, time-series swings in the average racial income ratio (over all age groups) in each census year merit a brief discussion. These are described in the final two rows of Table 12. According to my index, the relative income status of black men rose from 1890 to 1920, remained relatively constant between 1920 and 1940, increased sharply during the 1940's, and ended with the steady and significant post-1960 rise in relative black incomes.[16] Table 12 indicates not only that black-white income ratios rose after 1960, but also that the 1960 income ratio is clearly larger than that which prevailed throughout the 1920–40 period, and that the 1920–40 ratios exceeded the ratio of 1890. Affirmative action pressures of the 1960's cannot have been the exclusive driving force of black progress; the historical antecedents for some of the trends must have occurred earlier.

More germane to the issue addressed in this paper are the cohort-specific income ratios and the differential experiences of cohorts over their respective lifetimes. To highlight the central point, the generations

[14]At each stage in this process when additional aggregation over occupation groups was required, checks were made to test the sensitivity of the conclusions to the level of aggregation. I found no evidence that the level of aggregation into these 133 groups distorts the trends described in the text. For example, Table 12 was also computed for the post-1940 period using the more plentiful occupational groups available over this subperiod. The results were essentially identical to those using my 133 codes. The details of this process of making this series comparable are laborious, but are available from the author on request.

[15]Experiments indicate that it did not matter whether these income weights were based on the 1960 or 1970 Census. I also experimented with age groups 20–64, 14–75, and 5-year age groups. The results were insensitive to these experiments. In addition, comparisons with actual observed changes after 1940, when income data became available, indicate that at least over this period my index does not distort the actual changes that occurred. Clearly, my index understates the rate of post-1960 improvement by ignoring within-occupation narrowing of racial wage differentials.

[16]These decade-by-decade comparisons are necessarily made against the backdrop of major historical upheavals that can easily dwarf the type of slowly evolving forces that we argue were reshaping black America. For example, the Great Depression could easily have overwhelmed the factors at issue in this paper, for blacks surely bore more than their proportional share of its burdens. Thus, the 1940–30 comparison in Table 12 could easily suppress the slow economic evolution emphasized here.

TABLE 12—ESTIMATED BLACK-WHITE MALE INCOME RATIOS BY BIRTH COHORTS

Birth Cohort	Census Year 1890	1900	1910	1920	1930	1940	1950	1960	1970	1980
1956–60										.653
1951–55										.625
1946–50									.646	.624
1941–45									.608	.610
1936–40								.571	.583	.573
1931–35								.561	.582	.593
1926–30							.555	.555	.577	.583
1921–25							.542	.546	.571	.554
1916–20						.501	.532	.537	.562	.583
1911–15						.488	.516	.536	.562	
1906–10					.509	.480		.526	.548	
1901–05					.500	.482	.508	.524		
1896–1900				.515	.498	.483		.519		
1891–95					.494	.478	.503			
1886–90					.492	.469	.501			
				.494						
1881–85		.488			.481	.465				
1876–80					.458	.445				
1871–75	.472	.463		.387	.452					
1866–70					.438					
1856–65	.454	.451								
1846–55	.435	.428								
1836–45	.417	.398								
1826–35	.397									
All Ages 10–75	.433	.442	.455	.476	.479	.474	.516	.542	.590	.611
Ages 20–64	.439	.449		.484	.486	.479	.522	.543	.585	.605

born within the 20-year period of education retrogression are separated from the others in Table 12. To facilitate comparisons, Table 13 lists 20-year growth in average schooling by race. The cohort-specific patterns in these tables are remarkable. Both of the two historical periods of clear racial convergence in education—the first 25 years after slavery and the twentieth century—show steady across-cohort improvement in the relative income of black men. In contrast, the subperiod in which no strong across-cohort trends exist in relative income of blacks were those born within the era of education retrogressions. Using the 1930, 1940, and 1950 Censuses, the changes within the period of education retrogression average less than 1 percentage point. Table 12 also demonstrates that the rate of improvement for blacks accelerated as we move towards younger cohorts. This acceleration makes sense in light of the speedup in racial convergence in

TABLE 13—CHANGES IN AVERAGE EDUCATION LEVELS

Birth Cohorts	Black Males	White Males	Difference
1866–85	1.66	0.69	0.97
1886–1905	1.54	1.99	−0.45
1906–25	2.55	1.65	0.89
1926–45	2.81	1.31	1.50

education levels during this century (see Table 13), and of the point emphasized above that a substantial lag occurs before family background effects become operative.

The generations born between 1886 and 1905 have an effect on our interpretation of time-series changes in the relative income of blacks that lasts for half a century. These generations were the male labor force aged 20–34 in 1920, 25–44 in 1930, 35–54 in 1940, and 45–64 in 1950. Thus, they would dominate census surveys for close to 40 years. As long as they remained a large part of the total labor force, there was no reason to

expect, on the basis of human capital factors alone, much convergence in income ratios. Indeed, it is not until the 1960–70 census comparisons that these generations leave the labor market. And this is the census comparison for which the economics profession first took notice of improvement in economic conditions for black men.

C. *Labor Market Careers*

To this point, I have emphasized across-cohort trends and ignored within-cohort evolution over life cycle careers. The empirical origin of theories that blacks do not progress as rapidly as whites over their careers probably dates back to the 1940 Census, which first showed that black-white male income disparities increase systematically with age. This pattern has become a feature common to all cross-sectional studies of black-white earnings differences. It has been replicated in all the succeeding decennial censuses, and each of the 30-odd yearly *Current Population Surveys* beginning in the late 1940's.

The observation is well known that, by themselves, individual cross sections cannot test theories of life cycle discrimination. However, if we link these cross sections—for example, by reading across rows in Table 12—we can follow the actual experience of birth cohorts over their labor market careers. If we consider men who entered the labor force during the twentieth century (for example, those born after 1886), black-white income ratios did not decline over life cycles. The cross-sectional decline apparently results from more rapid cohort effects for blacks, and not from differential life cycle career paths. Among men born before 1886, Table 12 does provide evidence of a less rapid career income growth for black men, although obviously not as pronounced as the cross section would predict.

By construction, the ratios presented in Table 12 can only detect life cycle decay in the relative economic status of blacks that have their origin in movement across occupations. This table does demonstrate the important point that any cross-sectional decline in relative incomes of blacks due to different occupational distributions for men of different ages is a cohort and not a life cycle phenomenon. However, within occupations, incomes of blacks may be falling relative to whites, or more white-intensive occupations may exhibit faster career wage growth, even if race income neutrality prevails within occupations. If either of these two events occurred, relative income ratios of blacks would decline with age, but we would not capture it in Table 12.

Because of the absence of income data by race, it is not possible to observe actual income ratios over the long historical time span covered in Table 12. However, for more recent periods we can calculate series that are based on actual incomes. Tables 14 and 15 present my estimates of mean income ratios over the 1940–80 census years.[17] As one would expect, these actual income ratios are more erratic than those imputed using the more stable occupational distributions. While not without exception, the patterns depicted using ratios of actual incomes are consistent with the principal conclusions I have advanced using Table 12. Although each cross-sectional series in Tables 14 and 15 indicates a sharp decline with age, in the vast majority of cases black-white male income ratios either remained stable or actually rose over careers.[18] It is also worth noting that over the subperiod in which these two tables overlap, actual income series confirm trends in the imputed income series. Most important, the time period of retrogression in

[17]For the 1960–80 period, the actual income ratios are derived from the census and *CPS* micro data files. I calculated the 1940 series from tables of Zeman, which listed mean earnings by schooling and region. The 1950 series was derived from published tables listing income distributions by race, age, and schooling. Over the 1960–80 period, I have also calculated weekly wage ratios. The results with weekly wages were essentially identical to those with income. To preserve comparability with the 1940 and 1950 Censuses, only the income series is presented in the text. Because of small sample sizes, Table 15 stratifies schooling by 10-year cohort groups.

[18]The two main exceptions are 1921–25 and 1911–15 cohorts, principally during the first 10 years. There is a recovery during the remainder of the life cycle, but it takes most of it to achieve it. It is impossible to know for certain whether these are important exceptions or whether the income ratios are suspect.

TABLE 14—BLACK-WHITE INCOME RATIOS, 1940–80

Birth Cohort	Census Year 1940[a]	1950[b]	1960[c]	1970[c]	1980[d]
1950–60					74.9
1951–55					76.0
1946–50				82.6	76.1
1941–45				71.8	70.0
1936–40			69.1	64.7	69.0
1931–35			60.7	60.5	62.6
1926–30			54.9	57.2	62.7
1921–25		60.5	53.6	55.3	62.3
1916–20		55.0	50.4	54.1	67.4
1911–15	57.4	49.8	48.9	54.3	
1906–10	45.4		47.2	55.5	
1901–05	41.7	47.3	48.0		
1896–1900			45.9		
1891–95	42.5	46.6			
1886–90					
1881–85	42.9				
1876–80					

[a]Derived from tables in Zeman, using 1940 Census.
[b]1950 U.S. Census Special Reports, "Education," Table 12.
[c]1 in 100 decennial U.S. Census (computer tape).
[d]1980 *Current Population Survey* (computer tape).

TABLE 15—BLACK-WHITE MALE INCOME RATIOS 1940–80, BY YEAR OF SCHOOLING

Birth Cohort	Census Year 1940	1950	1960	1970	1980
Years of Schooling = 8–11					
1946–55					73.2
1936–45				72.8	76.4
1926–35			65.9	69.7	75.1
1916–25		68.4	68.6	69.9	82.6
1906–15	57.8	64.0	63.8	72.0	
1896–1905	51.8	62.5	61.0		
1886–95	51.6	63.0			
1876–85	51.7				
Years of Schooling = 12					
1946–55					75.1
1936–45				74.6	79.0
1926–35			65.1	70.3	72.0
1916–25		68.7	63.8	64.9	78.6
1906–15	55.3	61.6	57.8	65.0	
1896–1905	49.0	54.0	49.0		
1886–95	47.1	52.2			
1876–85	44.0				
Years of Schooling = 16+					
1946–55					88.5
1936–45				76.7	82.9
1926–35			62.5	78.0	79.1
1916–25		63.4	52.9	59.5	70.9
1906–15	49.6	53.4	45.7	54.9	
1896–1905	42.6	50.9	43.7		
1886–95	40.2	48.4			
1876–85	40.6				

education is a time of little increase in relative black male income in Table 14. Table 15, which presents these relative income series stratified by education level, indicates that school-quality trends are consistent with secular changes in levels. Within each schooling class, general trends do not support career retrogression for black men.

Given the initial stocks of human capital with which blacks and whites start their careers, black men of recent birth cohorts apparently do not face very different career prospects from those of white men. The fundamental problem is the wide disparities that exist when market competition begins, not what evolves over the labor market career. This conclusion is not a denial of the importance of labor market discrimination, but only about the form that it takes. The still large initial income differences between the races at the beginning of the career may well owe much to racial discrimination.

IV. Conclusion

During this century, America has undergone an evolutionary process, spawned in large part in its schools, that permanently altered and improved the relative economic position of black Americans. Because this process evolved slowly and because its implications for economic welfare were often not realized until decades after the seeds were planted, the importance and even existence of this evolution have been largely ignored or dismissed. But the historical evidence suggests that we have witnessed in this century a gradual but significant improvement in the quality and quantity of skills that blacks are able to carry with them into the labor market to compete with white workers. And as the human capital of blacks increased relative to whites, so also did their earnings.

REFERENCES

Ashenfelter, Orley, "Comment on Smith-Welch, 'Black/White Male Earnings and Employment: 1960–1970,'" in Thomas Juster, ed., *The Distribution of Economic Well-Being*, National Bureau of Economic Research, Studies in Income and Wealth, No. 41, Cambridge: Ballinger 1977.

Batchelder, Alan, "Decline in the Relative Income of Negro Men," *Quarterly Journal of Economics*, November 1964, *78*, 525–48.

Becker, Gary, *The Economics of Discrimination*, Chicago: University of Chicago Press, 1957.

Bond, Horace Mann, *The Education of the Negro in the American Social Order*, New York: Prentice Hall, 1934.

Darity, William A., "The Human Capital Approach to Black-White Earnings Inequality: Some Unsettled Questions," *Journal of Human Resources*, Winter 1982, *18*, 72–93.

Freeman, Richard B., "Changes in the Labor Market for Black Americans 1948–1972," *Brookings Papers on Economic Activity*, 1: 1973, 67–120.

Hanoch, Giora, "Personal Earnings and Investments in Schooling," unpublished doctoral dissertation, University of Chicago, 1965.

Hauser, Robert and Featherman, David, *Opportunity and Change*, New York: Academic Press, 1978.

Levin, Henry M., "Education and Earnings of Blacks and the *Brown* Decision," unpublished manuscript, September 1978.

Miller, Herman P., *Income of the American People*, New York: Wiley & Sons, 1955.

Myrdal, Gunnar, *An American Dilemma*, New York: Harper and Brothers, 1944.

Smith, James P. and Welch, Finis, "Black-White Male Wage Ratios: 1960-1970," *American Economic Review*, June 1977, *67*, 323–38.

Thurow, Lester, *Poverty and Discrimination*, Washington: The Brookings Institution, 1969.

Welch, Finis, "Education and Racial Discrimination," in Orley Ashenfelter and Albert Rees, eds., *Discrimination in Labor Markets*, Princeton: Princeton University Press, 1974.

Zeman, Morton, "A Comparative Analysis of White-Non White Income Differentials," unpublished dissertation, University of Chicago, September 1955.

U.S. Bureau of the Census, Special Reports of U.S. Census, *Current Population Reports*, 1947, P-20, No. 20, September 1948; 1950, P-E, No. 5B.

______, *Consumer Income*, Series P-60, various issues.

Competitive Payments Systems and the Unit of Account

By LAWRENCE H. WHITE*

Recent competitive innovations in payment mechanisms, particularly the checkable money market mutual fund, seem to have blurred the edges of the category of assets properly called "money." These innovations have coincided with new attempts by economists to reconstruct monetary theory and policy using competitive models. Several authors have conceived of competitive payments systems seemingly devoid of any outside currency, base money, or standard medium of exchange.[1] The unit of account in these systems is evidently not a common currency unit established outside the banking industry. Yet it can be argued that the use of a common unit of account in decentralized economic calculation presupposes a general medium of exchange.

Lance Girton and Don Roper have recently written: "One observes that most contractual obligations are specified in terms of the units in which the medium of exchange is measured. Further research should provide more insight into why contracts are specified in units in which the medium of exchange is measured" (1981, p. 20). This paper attempts to provide some insight into this question. By examining whether the above-mentioned cashless competitive payments systems are coherent and operational, it explores the fundamental relationship of the unit of account to the medium of exchange. It specifically examines the plausibility of competition divorcing the unit in which prices are specified (the unit of account) from the medium in which payment is typically made. The argument concludes that a payments system not based on convertibility into an outside currency should not be expected to arise in the absence of government intervention.

I. Cashless Competitive Payments Systems: A Brief Survey

A. *Black*

The belief that unrestricted competition would produce a payments mechanism devoid of outside money is expressed already in the title of Fischer Black's 1970 article, "Banking and Interest Rates in a World Without Money: The Effects of Uncontrolled Banking." Black claims that in the world he imagines "money in the usual sense would not exist" (p. 9). Initially he assumes that no currency is used; later he allows for currency, but supposes that its nominal quantity will be purely demand-determined, so that it does not serve as an outside money forming a base for bank liabilities.[2] Payments are made by transfer of this currency and bank liabilities. No mention is made of the redeemability of bank liabilities for this currency or any basic physical monetary asset produced outside the banking industry. I will for brevity's sake refer to such an asset as "outside currency" or "cash."

What serves as the unit of account? Black cannot say "the currency unit," for that is supposed to be subsidiary to the unit in which bank liabilities are denominated. Instead he says: "Goods may be priced in

*Assistant Professor of Economics, New York University, 269 Mercer St., New York, NY 10003. I am indebted for discussion and comments to Robert Greenfield, Leland Yeager, Fischer Black, Joseph Salerno, members of the colloquium on Austrian economics at New York University, and an anonymous referee. Research support from the U.S. Choice in Currency Commission (a private foundation) is gratefully acknowledged. Responsibility for the views expressed is mine alone.

[1]Fischer Black (1970), Eugene Fama (1980; 1982), Robert Hall (1981; 1982a,b), Robert Greenfield and Leland Yeager (1983). At the other extreme, F. A. Hayek (1978) and Benjamin Klein (1974) have conceived of a great multiplicity of parallel base monies and standards. Criticism of the latter models is left implicit in what follows.

[2]Currency in this world is supposed to be issued by the government, but only on request of the banks, in exchange for reduction of government debt with the banks. For criticism, see fn. 16 below.

terms of a unit of account that does not fluctuate in value very much, and means of payment may be priced in terms of the same unit of account" (1970, p. 14). The unit of account in Black's world is clearly not an outside currency unit as it is in our world. It is instead apparently a unit of a distinct numeraire commodity (or bundle of commodities) that does not itself serve as the means of payment. This is indicated by the remark that the means of payment is to be *priced* in terms of the unit of account rather than the unit of account being *defined* in terms of the means of payment. Thus Black's system divorces the unit of account from the characteristic units of the system's exchange media.

It is not at all clear in terms of what numeraire commodity the unit of account would be defined in Black's world, or how that numeraire would be selected. He conducts a thought experiment in which the means of payment successively assumes five forms: 1) barter; 2) shares of common stock; 3) corporate bonds; 4) corporate bonds certified by "banks"; 5) pure bank liabilities. The passage quoted above appears in his discussion of the second stage. There it was clear that the hypothesized unit of account was not the characteristic unit of the hypothesized means of payment (a share of a stock portfolio). At no later state is this divorce mended.

The logic of Black's construction receives fuller criticism below. But the following curious feature of Black's exposition deserves mention here. He speaks of "the dollar price" of a medium-of-exchange unit and "the dollar price" of a commodity, with "the dollar" clearly intended to designate the unit of account. He suggests that transactors in his system may use these "dollar" prices for the purpose of computing a commodity's price in terms of the medium of exchange. Yet there is nothing called "dollars" actually being traded against the commodities in the system, hence no mechanism for registering the prices of those commodities in terms of dollars. There are no dollar prices established on markets logically or temporally prior to establishment of medium-of-exchange prices.[3] The problem here is not that the unit of account is divorced from the medium of exchange, but that it is totally abstract, divorced from any traded good. Such an abstract unit of account, as Don Patinkin indicates (1965, p. 16), can have no operational significance for market participants. It can be meaningful only to a Walrasian auctioneer or other outside observer.

B. *Fama*

Black's article went uncited in the literature for a decade, until the appearance of Eugene Fama's "Banking in the Theory of Finance."[4] Fama, like Black, considers outside money inessential to the competitive payments mechanism he hypothesizes. He posits a "pure accounting system of exchange" (p. 42) in which the function of banks is to operate "a system of accounts in which transfers of wealth are carried out with bookkeeping entries" (p. 39). This method of wealth transfer is asserted to be "entirely different" in relevant respects from the use of cash. Fama claims that the transactions industry in the world he examines can dispense entirely with cash: "An accounting system works through bookkeeping entries, debits and credits, which do not require any physical medium or the concept of money" (p. 39). This means that in Fama's world, as in Black's, bank liabilities need not constitute claims to cash: "In a pure accounting system of exchange, the notion of a physical medium or temporary abode of purchasing power disappears" (p. 42).

Unlike Black, Fama is explicit in stipulating that the unit of account in his world should be thought of as the unit of a commodity that plays no medium-of-exchange role: "it could well be tons of fresh cut beef or barrels of crude oil" (p. 43). He explicitly recognizes that bank "deposits"—which would be heterogeneous, being essen-

[3] I am indebted to Robert Greenfield for this point.

[4] This result of a search through the literature (by Fama) was personally reported to me by Bob Hall. It evidently excludes self-citations by Black.

tially shares in various mutual funds and not claims to a common currency—are not a suitable candidate for numeraire.

Prices of commodities are stated in terms of the numeraire. Fama recognizes that an economy of this sort "is basically non-monetary." There is no question of price-level determination: since there is no money commodity trading against other goods, there is no money price level. There are only numeraire or relative prices to be determined. The determination of relative prices is apparently to be thought of as a performance of the Walrasian auctioneer. Fama speaks of the system posing "a standard problem concerning the existence of a stable general equilibrium in a non-monetary system" (1980, p. 44).

Like Black, Fama leaves the particular numeraire commodity ("some real good") and its method of selection both unspecified. This is of no concern so long as we take the auctioneer construct seriously. The auctioneer's choice of a numeraire is of no consequence. But Fama implicitly slips out of this construct. He suggests that agents in his world face genuine calculational problems, and that they deal with one another in decentralized markets rather than with the auctioneer alone. He says of the accounting system of exchange, for instance, that "its efficiency is improved when all prices are stated in units of a common numeraire" (1980, p. 43).

After analyzing banking in a nonmonetary setting, Fama introduces currency in the form of "a non-interest-bearing fiat currency produced monopolistically by the government" (1980, p. 50). The unit in which currency is measured may then serve as the economy's numeraire. The real value of a currency unit in terms of goods and services is determined in familiar fashion, as a determinate demand for real currency balances confronts a fixed nominal stock of currency.

Fama suggests that banks in the world with currency provide a "currency convertibility service" for their customers. But it is unclear whether he means "convertibility" in the usual sense of an obligation to *redeem* deposits on demand for outside currency. Banks taking on such an obligation have an inventory demand to hold currency as reserves against stochastic redemption outflows.[5] Limitation of the quantity of reserve currency available to the banks then limits the quantity of deposits that banks can prudently create. Fama states that banks would indeed "inventory currency on behalf of depositors" (1980, p. 50), but at the same time implicitly denies that the banks of an unregulated system would hold any non-interest-bearing reserves. Yet a bank's vault cash should be considered the primary component of its reserves where its deposits are convertible in the usual sense of constituting sight claims to predetermined quantities of currency.[6] By "convertibility," Fama must mean only that the banks act in the manner of money market mutual funds. Bank liabilities in his analysis are not claims to outside currency, as they are today, but are on the order of shares in a mutual fund's portfolio of interest-bearing assets. These funds (or Fama's "banks") stand ready to liquidate their shares (his "deposits") on demand by selling the assets to which the deposits constitute a claim and then turning over the proceeds to the shareholder ("depositor"). Fama is explicit in a more recent paper that this is what he envisions. He states that in his world: "Deposits are just claims against other claims (securities, loans, etc.)" (1982, p. 6). That is, they are not redeemable claims to outside currency. Fama's propositions that "deposits issued competitively should not be called money" and that "the concept of

[5]See Ernst Baltensperger (1980, pp. 4–6). In the competitive banking system of Scotland prior to 1844, to give a historical example, banks held positive quantities of specie as reserves against redemptions of liabilities despite the absence of reserve requirements and despite the fact, consistent with Fama's hypothesis of how a competitive system would operate, that the banks settled claims among themselves by transfer of readily marketable interest-bearing assets, namely Exchequer bills. On this episode, see my 1984 book, ch. 2.

[6]By "predetermined" I do not mean that deposit interest rates never vary, but that rates are contractually set before the period to which they apply. They are not calculated afterwards based on portfolio performance, as in the case of mutual fund shares. For further discussion, see Section IIIA below.

money plays no role in the transactions services accessed through deposits" (1982, p. 7) both rest on deposits not being claims to outside currency. The significance of the difference between such assets and deposits in the usual sense is explored below.

It is clear from the "parable" with which Fama concludes his earlier article (1980, pp. 55–56) that he regards the existence of outside money as unnecessary for the operation of an accounting system of exchange. Outside money is to him simply one commodity that, if it exists, may serve as numeraire; however, there is no need for it to exist. Steel ingots or spaceship permits may as well serve as numeraire. This result is arguably not true of any plausible world. There are compelling reasons, discussed below, for outside money to exist and to serve as the unit of account.

C. *Hall*

In two recent papers Robert Hall, searching for monetary policies consistent with stable prices and full deregulation of banking and financial markets, has questioned the necessity and desirability of associating the unit of account with a medium-of-exchange currency unit. Citing Fama (1980), Hall states: "It is possible to define the monetary unit [the unit of account] as one unit of a resource called currency, but this is only one of many different definitions" (1981, p. 4). In general the unit is simply "a certain amount of some resource" specified by government; the resource need not be currency. As an example of a noncurrency monetary unit, Hall proposes "defining" the dollar in terms of a composite-commodity unit called the *ANCAP*, consisting of specified physical quantities of ammonium nitrate, copper, aluminum, and plywood. Beyond defining the dollar in such a way, government is to play no role in the payments industry.

The *ANCAP* unit was chosen by Hall for its stable purchasing power over the last thirty years. Presumably this stability was measured in terms of some price index. An obvious question therefore arises: why does Hall not suggest defining the dollar directly in terms of the commodity bundle making up the price index he desires to stabilize? The answer lies in the mechanism he implicitly relies on for tying the value of the unit-of-account dollar to the specified commodity bundle. Only the commodity bundle is to be legal tender for dollar obligations. This means that all holders of contractual claims to receive dollars (or of obligations to pay dollars) are entitled to demand (or make) payment in the physical commodities defining the dollar. Any sufficiently wide divergence between the market price of the standard commodity bundle and one dollar will trigger demands by creditors to receive commodities rather than paper dollars (or deliveries by debtors of commodities in place of paper dollars). Transactors choosing to contract in *ANCAP* dollars would be exposing themselves to the risk of being forced to deliver, or to accept delivery of, physical bundles of the standard commodities. Every transactor would be taking on bank-like obligations. It is natural to doubt that many transactors would voluntarily do so. An *ANCAP* obligation seems to be clearly dominated for both creditor and debtor by an obligation indexed to the *ANCAP* bundle but contractually payable in a common medium of exchange, that is, explicitly ruling out the commodity-delivery possibility, given that a common medium of exchange is by definition more readily accepted than other commodities. The creditor would rather receive, and the debtor rather pay, readily spendable money than a bundle of commodities of equal market value. It is less implausible to suppose that specialized bank-like institutions might issue *ANCAP*-redeemable obligations. The question that then arises, to be answered below, is whether such obligations would gain currency in an unregulated environment.

D. *Greenfield and Yeager*

A recent paper by Robert Greenfield and Leland Yeager attempts to elaborate more explicitly the possible operation of a competitive mutual-funds-type payments system devoid of outside money. They attribute the inspiration behind the cashless competitive payments system to the three authors whose works I have just surveyed. In Greenfield and Yeager's view of that world, bank-like mutual funds would develop and operate a

sophisticated barter system (pp. 305–08). The unit of account would be an arbitrarily chosen numeraire bundle of commodities; the means of payment would be primarily shares of ownership in mutual fund portfolios. They explicitly affirm both the nonexistence of any outside money in which funds' liabilities are redeemable and the divorce of the unit of account from these media of exchange.

Greenfield and Yeager do not examine the question of whether such a system could emerge or survive under competitive conditions. They do consider whether the system's unit of account "has operational meaning" and whether "the level of prices expressed in that unit is determinate" (p. 313). In both cases, they find in the affirmative. But this merely means that they find the concept of keeping track of relative prices by use of a numeraire unit not incoherent or self-contradictory. It remains to be considered whether economic agents in an unregulated world without a central auctioneer would be likely to converge on use of a unit of account that is not a unit of outside currency.

II. Competitive Payments Systems in Evolutionary Perspective

In past and present monetary systems of our world, the generally accepted media of exchange have been and are units of outside money and inside-money claims to outside money. Inside money is naturally denominated in units of the cash to which it is a claim, as each banknote or bank deposit is a claim to a particular number of units of outside money. The distinguishing feature of outside money is that it does not constitute a redeemable claim to any physical asset. Whatever may be the bookkeeping conventions with regard to the issue of fiat money, as a form of outside money it is not in actual fact a contractual debt liability of any agent or institution. The world has known both commodity outside money—gold and silver coins provide the most familiar example—and fiat outside money. The latter typically originated as monopoly issued inside money whose redeemability was suspended after it had gained currency. In all cases the outside monetary unit naturally functions as the unit of account. This is because prices are naturally quoted in the units of the solitary item (or set of items, identically denominated because secondary members of the set are claims to a primary member) whose payment will routinely be accepted in exchange.

To mount a critique of cashless payments systems, one must give reasons for the emergence and prevalence of outside money as a generally accepted medium of exchange and unit of account. The reasons given here delve back to the origins of money.

A. *The Origin of Commodity Money*

The classic invisible hand explanation of the emergence of money from an initial state of barter was give by Carl Menger (1982). Under barter, each agent, attempting to transform his initial endowment into his desired final consumption bundle through direct exchange, confronts the problem of finding a second agent who both offers for sale what the first wishes to buy and is willing to accept in payment what the first has to sell. The typical agent can achieve his goal more economically if, instead of searching for this rare or even nonexistent match, he exchanges his endowment for more widely acceptable commodities that he may in turn readily exchange for the goods he ultimately wishes to consume. Accordingly he accumulates a trading inventory of highly saleable items. These allow him to economize on search costs by raising the probability that he may, in any given number of samplings among sellers, make desired purchases. In this situation the superior saleability of certain items becomes self-reinforcing: the knowledge that other traders will accept an item with high probability raises its acceptability to each particular trader. A network of traders will therefore converge on one or a small number of items as general media of exchange. Their supreme saleability then distinguishes these items from all other commodities. They have spontaneously become money.[7] Historically gold and silver emerged as money in eco-

[7]For a modern version of this theory, see Robert Jones (1976). See also Ludwig von Mises (1971, pp. 30–34). Menger defines "saleability" more or less as the narrowness of the effective bid-ask spread, but construes this broadly to include spatial and temporal dimensions.

nomically advanced nations through this process.

It should be readily apparent by extension of this perspective on the origin of money that a unit of account emerges together with and wedded to a medium of exchange. A seller pursues his self-interest by posting prices in terms of the media of exchange he is routinely willing to accept. This practice economizes on time spent in negotiation over what commodities are acceptable in payment and at what rate of exchange. More importantly, it economizes on the information necessary for the buyer's and the seller's economic calculation. Posting prices in terms of a numeraire commodity not routinely accepted in payment, by contrast, would force buyer and seller to know and agree upon the numeraire price of the payment media due. This numeraire price of the payment medium would naturally be subject to fluctuation, so that updated information would be necessary. A non-exchange-medium numeraire commodity would furthermore be subject to greater bid-ask spreads in barter against other commodities, as by hypothesis it is less saleable, than the medium of exchange. It would therefore serve less well as a tool of economic calculation.

It is worth emphasizing, as Menger emphasized with respect to the genesis of a general medium of exchange, that a collective decision is in no way necessary for the emergence of a clearly defined common unit of account. This point seems to have escaped those authors who consider monetary units to be the creatures of government proclamations.

B. *Coinage*

The evolution of monetary institutions does not, of course, stop with the emergence of commodity money. One may trace out further steps that take place in an unregulated competitive environment. Supposing that gold has emerged as primary money, the next logical step is economization of the costs of using the metal in transactions accomplished by the institution of coinage. Coined metal enjoys greater acceptability than uncoined metal (for example, gold dust) due to the lower cost of determining its true bullion content. The ease of authentication is still further enhanced by the institution of brand names in minting: once a mint's products are trusted to be of the weight and fineness stated on their face, its coins may pass by tale. Transactors may then forego weighing and assaying each piece of metal tendered in payment. The demand for readily authenticated pieces of gold will therefore give rise to a market in minting services. Each mint strives to maintain a reputation for uniformly high quality, lest it lose customers to its rivals by imposing higher authentication costs.[8] In competitive equilibrium, the mintage fee will be just sufficient to earn each minter the normal rate of return on investment. Self-interest will lead all mints in an economy to denominate coins in terms of a unit of standard weight and fineness. A mint doing otherwise would inconvenience its customers. The precise definition of the unit is itself unimportant; it may be based on preexisting custom in measuring the bullion content of uncoined gold, or it may be adopted from the coinage of an early reputable mint. This unit then serves as the unit of account.

Competitive private minting industries have been comparatively rare historically. Governments have typically monopolized the supply of minting services. In a noncompetitive situation, where debased government-issued coins circulate, the bullion content of an earlier full-weight coin may continue to serve as unit of account though no existing coin measures up to that content. This is the phenomenon of "ghost money," which is sometimes misleadingly cited as an example of divorce between the unit of account and

[8]For examples of this process at work in the United States, where some three dozen private mints operated in the gold rush regions of the nineteenth century, see Donald Kagin (1981). Black (1972, p. 811) inaccurately identifies privately minted coins as a form of inside money. Armen Alchian's (1977) account of the selection of a commodity money relies solely on economization of authentication costs. In my view, this explains the emergence of standardized forms of money, but as far as the origin of money itself goes is subsidiary to economization of search costs through holding of highly saleable commodities. Easy authentication is simply one among several properties contributing to ready saleability.

the medium of exchange.[9] In fact, the unit of account and the medium of exchange both continue to be quantities of gold. The unit-of-account value of any particular coin in circulation is a question of its weight and fineness, not of variable market exchange rates. The unit of account and medium of exchange have not become distinct commodities, only distinct quantities of the same commodity. The informational difficulties posed by a non-payment-medium numeraire, whose exchange value may vary in terms of payment media, do not arise. The minor inconvenience that does arise may be attributed to the absence of competitive conditions. Under competitive conditions, a debasing mint would find that money users reject its products in favor of full-weight coins.

C. *Bank Liabilities*

The emergence of precious metals as money, and subsequently of coins as their common form, comes about in a free economy as the undesigned outcome of decentralized pursuit of self-interest. The genesis of inside monies may be similarly explained. Bank liabilities originate as claims to specie deposited with bankers (hence the term deposits; Fama's use of this term to denote money market fund shares is misleading). In medieval Italy the first bankers were money changers; in London they were goldsmiths.

Claims to specie assume a monetary character when bankers discover profit in the business of effecting the payments one depositor wishes to make to another by direct transfer of bank balances from the one to the other. Checks are today the common means of signalling the bank to perform a transfer of balances, but the emergence of paperless electronic means would do nothing to change the essential nature of the transaction. Banknotes—claims to bank specie transferable without bank intervention and payable to the bearer on demand—similarly emerge as a means of payment.[10] Banknotes naturally find the greatest acceptance when denominated as round multiples of the specie unit that has previously become the standard unit of account. Money users find each form of redeemable claim to bank specie more economical to use for many purposes than actual specie. Bankers are recompensed for providing these instruments by the interest they earn on assets corresponding to the fraction of their liabilities not matched by specie on their balance sheets, or (in the case of deposits) by direct fees for the transfer service. In an unregulated system, the banks pay competitive rates of interest on their deposits. Due to the costliness of doing so, they are unlikely to pay interest on their notes.[11]

An invisible hand process can be shown (see my book, pp. 19–22) to account for the emergence of an interbank clearinghouse in a competitive banking system. Briefly, each member of a pair of banks profitably enhances the moneyness of its notes and deposits relative to specie by agreeing to accept one another's notes and deposits at face value as tendered by customers for deposit or loan repayment. Mutual acceptance of liabilities is naturally accompanied by an arrangement for periodic settlement of the claims each bank collects against the other. The potential gains from these pairwise arrangements are not exhausted until all banking companies in a region belong to a single clearinghouse system.

Members of the clearinghouse will, in the absence of regulation, be able to economize on specie transhipments by settling balances partly through the transfer of highly marketable interest-bearing assets. Specie redeemability remains essential to the economical functioning of the mutual acceptance ar-

[9]On "ghost monies," see Carlo Cipolla (1956, ch. 4). The misleading claim that these represent abstract units of account is made by Patinkin (p. 15). While it is true that a ghost money unit had no exact counterpart among existing coins, each of these coins bore a fixed value relationship to the unit based on relative bullion content. For purposes of pricing and calculation, the situation was similar in kind to that prevailing today in the Italian monetary system, where no one-lira coin or note circulates.

[10]On the early history of European banking, see Raymond de Roover (1956, ch. 5).

[11]See my book (pp. 8–9). Fama (1982, pp. 14–15) comes to the same conclusion for currency that is not a claim to outside money. Note that today's traveler's checks do not bear interest.

rangement, however, as the means by which all bank liabilities have their value fixed in terms of the unit of account. The acceptance of their notes at fixed par values spares banks' customers—and the banks themselves—exchange risk and calculational inconvenience, and is therefore integral to the function of acceptance arrangements in enhancing the moneyness of the participating banks' liabilities.

A competitive banking system of the following sort thus emerges in the absence of regulation. The stock of exchange media consists of specie in the hands of the public plus numerous brands of redeemable banknotes plus transferable bank deposits. The self-interest of issuers insures that notes circulate at par, that is, at unit-of-account values fully equal to the number of specie units to which they are claims.[12] Transferable deposits bear a competitive rate of interest, subject to competitive charges for transfer services. The nominal quantities of specie, notes, and transferable deposits held by the public are determined not by any central bank regulation of the monetary base, but by the real demand to hold those assets divided by the purchasing power of specie. Each bank's holdings of specie reserves are determined by its equating at the margin the cost of foregone interest to the benefit of reduced risk of illiquidity. Total specie reserves are simply a summation of these holdings across banks.[13]

The transition from a specie-based competitive banking system to a fiat-currency-based system is most readily made in two steps: government creation of a central bank, whose specie-redeemable liabilities displace specie as a commercial bank reserve asset; and suspension of redeemability for central bank liabilities. The supply of banking services may continue to be competitive, but the nominal quantity of money is now scaled to central bank determination of the monetary base.

Note what happens to the unit of account in the transition to fiat money. At no point does it cease to be defined in units of the basic outside-money medium of exchange. The status of basic medium of exchange, however, passes from specie alone to a straddle between specie and a redeemable central bank currency denominated in specie units (dollars, pounds sterling, etc.), then to the no-longer-redeemable central bank currency (still bearing the same name) alone. In this way the economy arrives at a situation in which a noncommodity outside money has positive exchange value. Paper money is able to function as the basic medium of exchange because it previously functioned as a secondary medium of exchange.[14]

III. Cashless Competitive Payments Systems: Critique

In light of the evolution of money and banking, the problem confronting models of noncurrency-based payments systems is clear. Their applicability for modeling current institutions or predicting future arrangements awaits a coherent account of how a cashless system is consistent with or might emerge from the currency-based payments systems the world has known. This is not to deny that such models may serve to illuminate the monetary institutions of our world by contrast to the abstraction of a world without outside money. This is a use to which Greenfield and Yeager deliberately put their model. It is a role Fama may also have in mind, as he later introduces outside currency to his model after first abstracting from it. In a way, the models play this role in the present discussion: I hope to illuminate the importance of the causal-genetic processes behind monetary institutions, particularly the unit of account, by contrast to models seemingly inconsistent with these processes.

[12] That banknotes fell below par when they crossed state borders—reflecting risk and transportation costs of accomplishing redemption—in the American "free banking" era was due to the legal prohibition on interstate branch banking. In the freer Scottish system, no such inconvenience was experienced.

[13] This system is spelled out in my book (ch. 1). The statement of marginal conditions in the text assumes equal marginal operating costs of holding various assets. The basic paradigm of bank optimization is set forth by Baltensperger.

[14] This historical account may explain the fact that intrinsically useless fiat money has positive value more plausibly than the overlapping generations model of fiat money. For that model, see Neil Wallace (1980).

A. *The Disappearance of Demand Deposits*

Could a monetary system based on outside currency (specie or fiat currency) spontaneously evolve into a cashless competitive payments system of the sort envisioned by Black, Fama, and Greenfield-Yeager? Three steps are necessary to make the transition: 1) disappearance of redeemable inside money; 2) disappearance of outside money; and 3) redefinition of the unit of account in terms of a numeraire other than outside money. This section considers the first of these steps. For expositional convenience it focuses on demand deposits, though in the past banknotes have also been important as inside money. The term inside money here denotes ready claims to outside currency. These are distinct from shares in a managed portfolio of assets.

Fama envisions a world in which "competitive unregulated banks provide a wide variety of portfolios agains which depositors can hold claims" (1982, p. 15). Bank deposits no longer constitute claims to cash, in other words, but are instead akin to transferable shares in mutual funds and hence "can be tailored to have the characteristics of any form of marketable wealth" (Fama, 1980, p. 43). Fama unfortunately fails to show that the outcome of unregulated competition would be the total domination of interest-bearing demand deposits by mutual fund shares. In fact this outcome is unlikely, even apart from the question of which can provide payments services more efficiently. Demand deposits, being ready debt claims, are potentially superior to mutual fund shares, which are equity claims, in at least one respect. The value of a deposit may be contractually guaranteed to increase over time at a preannounced rate of interest. Its unit-of-account value at a future date is certain so long as the bank continues to honor its obligation to redeem its deposits on demand. No such contractual guarantee may be made with respect to an equity claim. A mutual fund is obligated to pay out after the fact its actual earnings, so that the yield on fund shares cannot be predetermined. In the absence of deposit rate ceiling regulation, the range of anticipated possible returns from holding fund shares need not lie entirely above the deposit interest rate. Risk-diversifying portfolio owners might therefore not divest themselves entirely of demand deposits even given a higher mean yield on mutual funds. It is true that the characteristic pledge of money market mutual funds to maintain a fixed share price, or rather the policy of investing exclusively in short-term highly reputable securities so that the pledge can be kept, makes fund shares akin to demand deposits in having near-zero risk of negative nominal yield over any period. The difference between predetermined and postdetermined yields—between debt and equity—nonetheless remains. The historical fact is that deposit banking did not naturally grow up on an equity basis.[15]

The more important reason why demand deposits would survive even under unregulated competition is that the payments system they provide is, given the conditions that lead to the emergence of money, less costly. This cost differential is suggested by the fact that a checkable money market fund today typically imposes a $500 minimum on checks written against shares in the fund. The comparative costliness of check writing against money market funds in their present form arises from the fact that checks written against a fund require it either 1) to incur the transactions costs of selling securities plus the cost of transmitting the receipts to the payee, or 2), what is presumably less costly and the method actually used, to draw against a demand deposit with a commercial bank held as one of the fund's assets.[16] In the latter case, it is evident that effecting a payment by writing a check against a fund, which in turn draws down its demand deposit, must be more costly than directly

[15]Though there was medieval banking in which bank deposits were treated as equity claims, this treatment was devised to evade church and state prohibitions against the payment of interest on debt. Again see de Roover (pp. 201–02).

[16]All funds whose prospectuses I have examined hold a small percentage of their assets (less than 1 percent) in the form of a demand deposit with a commercial bank for the purpose of honoring redemption checks (and purchasing securities).

effecting the payment by writing a check against the payer's own demand deposit. In the present world the checkable money market fund rides piggyback upon the banking system.

The check writing feature of money market mutual funds relies on a money-transfer system for the obvious reason that sellers of commodities generally wish to be paid in money and not in other assets. Checks written on a money market fund are generally acceptable in payment only because to the recipient they represent a transfer of inside money, that is, of cash-redeemable bank deposits. Its unique acceptability as a routine means of payment is, as we have seen, an essential property conferred on money by the Mengerian convergence process that engenders money. Every form of marketable wealth could serve generally as a medium of exchange only in a world where all forms of wealth begin and remain equally marketable. Outside a Walrasian general equilibrium setting, this is difficult to imagine.

There are no obstacles in principle to the spontaneous emergence of an interfund clearing system that does not rely on transfers of inside money. If mutual funds really could provide payments services efficiently, it would be natural to expect money market funds in the present system, unless prevented by law, to begin announcing bilateral or multilateral arrangements to permit check writing in any amount for purposes of transferring wealth to accounts in participating funds. By this device, each participating fund would enhance the spendability and hence desirability of its shares relative to nonparticipating shares and demand deposits. As yet this has not happened. At present, money market funds rarely allow check writing for unlimitedly small amounts, even for transfer of shares to another customer of the same fund. This is difficult to reconcile with the idea that fund shares are so routinely acceptable that they could dominate inside money as a means of payment.

This argument does not rule out mutual funds developing a money-transfer system and allowing cash withdrawals, or what would be identical, banks offering checkable mutual fund accounts with direct access to an interbank clearing mechanism. The analytical question in this case—why money-transfer and cash-inventory services should be jointly produced with deposits at lower cost than with mutual fund shares—awaits further research. But it seems clear that the major impetus to the use of mutual funds for check writing purposes, a use negligible before 1974, has been Regulation Q's prohibition of competitive interest rates on checkable bank deposits. With this ceiling largely lifted, the rationale for joining money-tranfer services to mutual funds has largely disappeared.[17]

In a model competitive payments system devoid of cash or genuine demand deposits, payments effected via check writing against fund shares obviously do not work by transfers of money. Instead a check written against Fund A in favor of a customer of Fund B is supposed to occasion a transfer of nonmonetary assets from Fund A to Fund B via a clearing arrangement (Greenfield-Yeager, p. 307). These two funds must have previously entered a mutual acceptance arrangement of the sort (described earlier) arising in a free banking system. The clearing mechanism has to be slightly different, however, in the following respect. Fund B, in accepting checks written on Fund A, does not possess a claim to Fund A's vault cash of a specific quantity. Instead Fund B possesses a claim to Fund A's assets of a specific value. Checks are written, and interbank clearing balances computed in units of account, as at present. But a check no longer transfers a claim to so many physical units of outside currency; it instead transfers ownership of earning assets with a market value of so much. The interfund clearing arrangement has to specify

[17]Two caveats are in order. 1) The 1982 Garn-St. Germain Act authorizing Super NOW accounts (checking accounts with no legal interest ceiling) denies these accounts to business firms, leaving firms a reason for using money market fund or sweep accounts for check writing. 2) So long as demand deposits are in effect taxed by the imposition of reserve requirements, there remains a rationale for hybrid accounts. The reason why money market mutual funds (like banks) do not price their money-transfer services explicitly may be found in the taxation of explicit interest but nontaxation of gratuitous services.

what types of assets are acceptable in settlement of adverse balances. So does an interbank clearing arrangement if it is to economize on physical transfers of non-interest bearing currency, of course, but this does not reduce its reliance on cash redeemability as the means by which the unit-of-account value of bank liabilities is fixed and their general acceptability maintained.

An apparent disadvantage of bank deposits in the form of ready claims to predetermined quantities of currency, in contrast to fund shares, is the possibility that a bank might become insolvent and thereby unable to honor all the claims presented to it for redemption. (Illiquidity is no greater problem for a bank than for a mutual fund that allows check writing and cash withdrawals.) A mutual fund cannot become insolvent: as it issues no liabilities in the strict sense, but only equities, it cannot have liabilities in excess of its assets. A money market fund can legally break its pledge to maintain a fixed share price if a sharp fall in the value of its assets makes a reduction necessary. A bank lacks the flexibility to reduce its deposit liabilities in a similar way without going into bankruptcy. In a *laissez-faire* monetary system, bank deposits would not be government insured. Depositor fears of insolvency might be adequately addressed, however, by high capital-asset ratios, by private deposit insurance, by forms of organization giving the bank's stockholders extended personal liability for its debts, or by some other means.[18] Hence it is not obvious that checkable mutual funds would dominate demand deposits on grounds of lesser risk. The debt form of deposits does insulate depositors from sharing in portfolio losses that leave equity positive.

The difference between demand deposits and fund shares, and the plausible nondisappearance of the former under freely competitive conditions, requires the revision of several propositions put forth by Fama (1982, pp. 2–8). 1) While outside currency and fund shares are indeed not perfect substitutes whose supplies may with any obvious sense be aggregated, and while outside currency and demand deposits are also not perfect substitutes, demand deposits (and banknotes) may sensibly be aggregated with outside currency held by the nonbank public in a measure of the quantity of money. The econometric use of this aggregate is a separate question. 2) The supply of demand deposits will likely be important in the determination of the price level for a closed economy with a competitive unregulated banking system. Even if the determination of the price level in that economy is most appropriately modeled in terms of the supply and demand for outside money alone, demand deposits are presumably a close substitute on the demand side. 3) The concept of money clearly does play a role in the transactions services made available through demand deposits. 4) A bank using the clearing mechanism of an unregulated banking system holds claims against the cash reserves of other banks, not against their portfolios.[19]

B. *The Disappearance of Outside Money*

Might outside money disappear with the evolution of competitive payments mechanisms? This boils down to the question of the disappearance of outside currency. In the present American banking system, the deposits of member banks with the Federal Reserve may be regarded as a form of outside money (though they are claims to Federal Reserve notes, their quantity is not regulated by the existing quantity of those notes). This form of outside money is an artifact of regulation, however; in an unregulated banking system with a private clearing mechanism and no central bank, outside currency (say, specie or fiat currency) would be the only form of outside money.

The authors whose models have been considered here all recognize that currency will continue in use so long as manual transfer of

[18] Unlimited liability was a feature of the Scottish free banking system. Depositors' losses due to bank insolvencies were completely negligible, as failures were rare and the losses fell upon shareholders.

[19] Only the last of these sentences rectifies an incorrect statement Fama makes about a banking system. The others contrast a banking system to his characterization of a payments system operated by mutual funds.

currency remains the least costly method for accomplishing certain transactions. Not only is currency 1) more convenient to use in small payments, but 2) its acceptance, unlike acceptance of personal checks, entails no risk that the payer's funds may be insufficient, and 3) its use leaves behind no possibly incriminating records of payment. These authors all think it coherent, however, to suppose that all currency is inside currency. Pieces of such currency would be akin to banknotes, except that they would constitute claims against the portfolios of the issuing funds rather than claims to cash.[20]

Cashlessness has an important implication. Bonds in the cashless world cannot be what they are in our world, claims to future streams of money payments. They must rather be claims to future payments of commodities or to other financial assets. These other financial assets must be equities or shares in a mutual fund portfolio of equities, as it would be circular for bonds to be exclusively claims to other bonds, either directly or indirectly via money market fund shares in bond portfolios. The present value of bonds in the cashless world must then be the discounted value of the commodities or equities to be received in future payments. This clearly would make bond pricing much more difficult than it is in our world were the future payments to be defined in units of the commodities or equities to be paid. Greenfield and Yeager understandably suggest that the quantity of payment property (as they call it, p. 313) to be received would be specified, like all other contractual payments, in numeraire value units rather than in the physical own-units of the property. Coupon payments would proceed in commodities or equities of specified worth in terms of the numeraire. The bondholder nonetheless receives payment in commodities or equities. In general he will wish to sell these rather than hold his wealth in their form, so that he will prefer bonds whose coupon payments are made in the most readily saleable form of property. In our world the most saleable property is money; in the cashless world it is supposed to be shares in mutual fund bond portfolios. But this, as I have noted, creates a circularity problem. Hence one of two outcomes is possible: either bondholders are saddled with relatively high transactions costs in unloading payment property, or bond portfolio shares are not the dominant means of payment. In the latter case, say where shares in a mutual fund portfolio of common stocks were instead the dominant means of payment, the numeraire value of exchange-medium holdings would clearly be subject to significant fluctuation.

The natural question to ask from an evolutionary perspective is whether there is any plausible reason for outside currency to disappear in a payments system freed from anticompetitive regulation. I have explained above that the emergence of particular commodities as money is not wholly accidental, but a consequence of their superior saleability. Black (1970, p. 14) hypothesizes the use of shares of a portfolio of common stock as money, that is, as a generally accepted medium of exchange. There are good reasons, however, to doubt that such an item would ever become the most saleable in an economy. The primary reason is that the institution of common stock is unlikely to arise in a premonetary economy because the division of labor it presupposes would not exist there. Even were stock shares to emerge in a barter economy, it is difficult to conceive of their being more saleable than the most widely saleable of commodities. Arising in an already monetized economy (this is Black's scenario), shares of stock are from the outset routinely sold against money and not against any other good. They lack the saleability of money. And this inferior saleability is self-reinforcing: no trader routinely accepts shares of stock or shares of a portfolio of stocks when he cannot expect to be able to spend them easily. Each trader finds the use

[20] Fama (1982, pp. 9–11) and Greenfield-Yeager (pp. 307–08) clearly envision currency issued exclusively by mutual funds. Black (1970, pp. 13–14) introduces government-issued currency, but erroneously believes that the nominal quantity of this currency will be endogenously determined. He apparently fails to see or denies that an excess of supply of government currency at a given level of prices will be worked off through a rise in prices, not through retirement of the excess currency. In another paper (1972), Black advances a doctrine of the passivity of outside money.

of shares an inefficient medium of exchange due to high information and search costs. The "inefficiencies" of commodity money cited by Black would exceed the inefficiencies of common stock money only in a world in which common stock approached the saleability of commodity money.

For analogous reasons it should be apparent that a commodity reserve currency system, in which the basic money is redeemable for a basket of nonmonetary commodities, would not arise spontaneously in an unregulated setting. A claim to a basket of commodities would not originally emerge as money, since in a barter setting it would be less saleable then the most saleable of its components. Nor would it supplant the original monetary commodity. This is not to deny, however, that one money (say, silver or domestic fiat currency) may be spontaneously supplanted by another (say, gold or foreign fiat currency) in a region where both have been circulating internally, or where external trade with neighboring regions is conducted in their different money. A switch may come about because the transactions conducted in the second money grow in relative importance, or because the first money experiences an exogenously caused ongoing relative decline in purchasing power.

C. *The Divorce of the Unit of Account from the Medium of Exchange*

For reasons already suggested, a unit of account emerges wedded to a general medium of exchange. Prices are universally posted in the characteristic units of a medium or set of media that sellers are routinely prepared to accept in exchange. This process is self-reinforcing: a buyer or seller who communicated bid or ask offers in nonstandard units would impose calculation costs on potential trading partners. For this reason the unit of account remains wedded to the medium of exchange.

In an inflationary environment it is certainly possible for a unit of stable purchasing power to dispace the depreciating currency unit as the unit of account voluntarily adopted in contracts calling for payments at future dates. An example of a stable unit would be the "constant dollar" defined by a base-year price index. There is no tendency for spot prices to be indexed in this way, however. Indeed the perpetuation of nonindexed spot prices is presupposed by indexing, which uses current nominal prices to compute the current-dollar equivalent of a constant-dollar sum.

The unit of account sticks with the medium of exchange even through the transition from commodity-based to fiat currency. A historical example is instructive here. In the suspension period of the Napoleonic Wars, 1797–1819 in Britain, Bank of England notes and deposits became the basic outside money.[21] Gold coins ceased to circulate. The unit of account, the pound sterling, stuck with the actual medium of exchange rather than with a now-abstract gold definition. The pounds-sterling price of gold fluctuated rather than the pounds-sterling price of Bank of England notes. Commodity prices rose with the expansion of Bank of England notes and deposits, while the unit-of-account value of a banknote or deposit remained fixed.

IV. Conclusion

In a decentralized and unregulated economy in which all property is not equally saleable, outside money emerges as most the saleable commodity and persists as a general medium of exchange. Inside monies arise and persist on the basis of their convertibility into outside money. The characteristic unit of outside money naturally defines the unit of account, as prices are natually posted by traders in terms of the item sellers will routinely accept in payment.

In a Walrasian world where the auctioneer renders all commodities equally saleable, and therefore equally suitable for use in indirect exchange, payment in any commodity could be accepted indifferently. Tatonnement may proceed without outside money. Any commodity or bundle of commodities could serve as unit of account, the auctioneer's choice of a unit of account being unconstrained by any economic considerations. The payments

[21]Technically they were not fiat money since resumption at a later data was both anticipated and realized. In von Mises' (p. 483) terminology they were credit money.

system appropriate for such a world, however, is inappropriate in the present world of decentralized trade involving goods of unequal marketability. The convenience of traders in the present world dictates outside money whose units define the unit of account. Deregulation of the payments system in the present world does not imply disappearance of outside money, nor divorce of the unit of account from the basic outside-money medium of exchange.

REFERENCES

Alchian, Armen, "Why Money?," *Journal of Money, Credit and Banking*, February 1977, *9*, 133–40.

Baltensperger, Ernst, "Alternative Approaches to the Theory of the Banking Firm," *Journal of Monetary Economics*, January 1980, *6*, 1–37.

Black, Fischer, "Banking and Interest Rates in a World Without Money: The Effects of Uncontrolled Banking," *Journal of Bank Research*, Autumn 1970, *1*, 9–20.

______, "Active and Passive Monetary Policy in a Neoclassical Model," *Journal of Finance*, September 1972, *27*, 801–14.

Cipolla, Carlo M., *Money, Prices, and Civilization in the Mediterranean World: Fifth to Seventeenth Century*, Princeton: Princeton University Press, 1956.

de Roover, Raymond, *Business, Banking, and Economic Thought in Late Medieval and Early Modern Europe*, Chicago: University of Chicago Press, 1956.

Fama, Eugene F., "Banking in the Theory of Finance," *Journal of Monetary Economics*, January 1980, *6*, 39–57.

______, "Fiduciary Currency and Commodity Standards," mimeo., January 1982.

Girton, Lance and Roper, Don, "Theory and Implications of Currency Substitution," *Journal of Money, Credit and Banking*, February 1981, *13*, 12–30.

Greenfield, Robert L. and Yeager, Leland B., "A Laissez Faire Approach to Monetary Stability," *Journal of Money, Credit and Banking*, August 1983, *15*, 302–15.

Hall, Robert E., "The Government and the Monetary Unit," mimeo., 1981.

______, (1982a) "Explorations in the Gold Standard and Related Policies for Stabilizing the Dollar," in his *Inflation: Causes and Effects*, Chicago: University of Chicago Press, 1982.

______, (1982b) "*Monetary Trends in the United States and the United Kingdom*: A Review from the Perspective of New Developments in Monetary Economics," *Journal of Economic Literature*, December 1982, *20*, 1552–56.

Hayek, F. A., *Denationalisation of Money*, 2d ed., London: Institute of Economic Affairs, 1978.

Jones, Robert A., "The Origin and Development of Media of Exchange," *Journal of Political Economy*, November 1976, *84*, 757–75.

Kagin, Donald H., *Private Gold Coins and Patterns of the United States*, New York: Arco Publishing, 1981.

Klein, Benjamin, "The Competitive Supply of Money," *Journal of Money, Credit and Banking*, November 1974, *6*, 423–53.

Menger, Carl, "On the Origin of Money," *Economic Journal*, June 1982, *2*, 239–55.

Patinkin, Don, *Money, Interest, and Prices*, 2d ed., New York: Harper & Row, 1965.

von Mises, Ludwig, *The Theory of Money and Credit*, rev. ed., Irvington-on-Hudson: Foundation for Economic Education, 1971.

Wallace, Neil, "The Overlapping Generations Model of Fiat Money," in John H. Kareken and Neil Wallace, eds., *Models of Monetary Economies*, Minneapolis: Federal Reserve Bank of Minneapolis, 1980.

White, Lawrence H., *Free Banking in Britain: Theory, Experience, and Debate, 1800–1845*, Cambridge: Cambridge University Press, 1984.

Reputations in the Labor Market

By H. LORNE CARMICHAEL*

For several years the dominant approach to theoretical work in labor economics has built upon the idea of an "implicit contract." This approach characterizes the relationship between a firm and its workers as a two-stage game. In the first, or *ex ante*, stage, there is no production but the two parties do get together and cooperate to lay down a set of rules that will govern their future interactions. Information is imperfect in that neither party knows the states of the world that will obtain in future periods. Information is, however, symmetric in that the distribution of states of the world, the firm's technology, and the worker's utility function are all public knowledge. In the second, or *ex post*, stage, production may take place if conditions warrant. However, each party is now assumed to behave noncooperatively, although neither breaks any of the rules laid down in stage one.

The early work in this area is associated with Costas Azariadis (1975), Martin Neil Baily (1974), and Donald Gordon (1974). This work assumes that *ex ante* agreements cover only wage and employment levels, and that information about the state of the world *ex post* is revealed to each party. Agreements made *ex ante* can therefore require that wage and employment levels be contingent on the state of the world. The major conclusions are that if workers are risk averse and firms are risk neutral, then workers will enjoy fixed real wages, but may experience unemployment in the second period. This unemployment will be involuntary, in that workers will prefer to continue working at their old wage, but it will also be efficient (in partial equilibrium) in that the value of a laid off worker's time at home will always be at least as great as the value of his time at work.[1]

More recently, research by Robert Hall and David Lilien (1979), Sanford Grossman and Oliver Hart (1981), and Azariadis (1983) has examined the form of optimal *ex ante* agreements when workers do not observe the state of the world *ex post*. Each party remains well informed *ex ante* about everything else. Optimal contracts in this setting tend to make the worker's wage a function of his hours of work (since these are observable). General conclusions are that wages will be monotonic in hours of work and that employment levels will not be set efficiently in the *ex post* sense, although the agreements are *ex ante* efficient given the informational structure. Whether the result is too much or too little employment seems to depend on the specification of the model (see Russell Cooper, 1983).

The basic structure of this approach seems to bear very little resemblance to the actual process by which institutions evolve in the nonunion sector of the labor market. Its relevance therefore depends on the extent to which firms and workers behave "as if" they were entering into explicit legally enforceable agreements. There are two aspects to this problem. First, the policies of nonunion firms are not usually the subject of bargaining. Rather, firms design policies, or regular patterns of behavior vis-à-vis wage profiles, layoffs, etc. in order to enhance their position in a competitive labor market. Is it clear that the outcome of this dynamic process is the same as the outcome of a one-shot *ex ante* bargaining game? Second, even if it is justifiable to make the *ex ante*, *ex post* distinction, it is typically the case in these mod-

*Queen's University, Kingston, Canada K7L 3N6. I thank the many people who expressed their dissatisfaction with previous versions of this paper, including Curtis Eaton, Russel Cooper, Jon Strand, Sherwin Rosen, participants of the NBER Summer Workshop on Economic Fluctuations, and an anomymous referee. Financial support from the Social Sciences and Humanities Research Council of Canada and the Hoover Institution is gratefully acknowledged.

[1] To avoid confusion, this is the definition of the word "efficiency" that will be used throughout the paper. Whether or not the free market or a benign central planner could achieve this result is another matter. See Section III.

els that one or both of the parties will find it in his interest to change the rules once the state of the world has been revealed. Since the "contracts" are not in writing and not enforceable in any legal sense, what is it that prevents the parties from cheating?

These questions have already received extensive treatment in the literature. The first is not usually thought to be a serious problem. With an explicit contract, the first-order conditions for a Pareto optimal outcome in the *ex ante* bargaining problem are the same as those from the firm's problem of maximizing profits subject to the competitive hiring constraint. The problem of preventing the worker from reneging has also been treated satisfactorily in a multiperiod framework by Bengt Holmstrom (1983). Enforcement on the firm's side has relied on the assumption that the firm cares about its reputation as an employer. A firm that arbitrarily changes the terms of an agreement with its workers will presumably have trouble hiring workers in the future.

In recent years there have been several papers examining the role of reputations in intertemporal decision problems. Roy Radner (1981), and Robert Townsend (1982) among others consider reputation formation in repeated games. Benjamin Klein and Keith Leffler (1981) and Carl Shapiro (1982) consider the role of reputation in the enforcement of product quality in the output market. Robert Barro and David Gordon (1983) consider reputations in a macro model. In the labor market, Edward Lazear (1981) and Oliver Hart (1983) briefly discuss reputations, while Holmstrom (1981) and Jon Strand (1983) provide explicit treatments.

An important implication of this literature seems to be that while reputations will not perfectly enforce an implicit contract, they do have an advantage in that they are not affected by the presence of asymmetric information. In a reputational model, workers base their expectations of future wage and employment levels on actual past values. Firms know this and take it into account when they set current levels. This process works independently of any information workers may or may not have about the current state of the world. Since contracts under asymmetric information will not normally provide first-best outcomes anyway, reputational mechanisms may be preferred in some circumstances.

This argument suggests that it may be useful to explicitly investigate the role of reputation in the labor market not simply to see if the assumptions made in the contract literature can be justified, but also, and more importantly, as a first step in the construction of an independent paradigm for the examination of labor market issues. This paper attempts to do just that. In the process it is able to shed light on previous work on implicit contracts, as well as on several institutional features of modern labor markets including wage rigidity, unemployment, severance pay, government regulation, and union growth.

This paper examines two models of a competitive labor market where employers do not attempt to enter explicit or implicit agreements with present or prospective workers. Rather, they establish regular patterns of behavior which they hope will enhance their competitive position in the market. Perhaps the strongest justification for this approach is the simple fact that explicit contracts are not observed in the nonunion sector of the labor market. Since the explicit modelling of reputation leads to different implications than those obtained by contract models (as will be made clear in this paper), an important justification for the use of contract models has been lost. In order to fully understand the role of reputation in the labor market, we must model the process by which reputations are formed.

The plan of the paper is as follows. The basic structure of the model is developed in Section I, and several results which bear on previous work in the contract paradigm are obtained. In particular, the earlier result that reputation by itself will not enforce efficient behavior on the firm is reproduced, and it is shown that even though there may exist institutions (in this case, an employment contingent wage contract) that if imposed exogenously would restore efficiency, the firms in a reputational market will not introduce them. Previous results from asymmetric information contract theory are therefore shown to

depend critically on the assumption that these "implicit contracts" are in fact explicit and legally enforceable. Their application should therefore be confined to the union sector of the labor market. Section II examines another model in order to generate some new empirical implications. The two main results from this section are that severance pay will typically be underprovided even though workers are risk averse and firms risk neutral, and that the rate of interest will have a direct, short-run effect on the level of employment in the economy. Section III discusses some of the implications of the paper concerning government regulation and union growth.

I. A Simple Reputation Model

This section builds a simple model of a reputational labor market and establishes most of the theoretical results. Of immediate interest is the actual method by which reputations will be modelled. In general, there are two aspects to what is meant by a firm's "reputation." In any model it is clearly defined by workers' expectations as to the firm's future behavior. If workers expect the firm to treat them poorly, then the firm has a "bad reputation." More than this, however, is the requirement that a reputation be based on the firm's actual behavior in the past, rather than upon promises or explicit guarantees. The most direct way of modelling these ideas is simply to make workers expectations adaptive on the firm's past behavior.

Adaptive expectations have a long history in macroeconomics. More recently, however, their use has been completely supplanted by techniques that require agents to have much more information about the structure of the economy. As a behavioral assumption, this movement to rational expectations can be justified in several ways. In financial markets, any extra information that agents can use to help predict the future movements of prices can quickly be transformed into extra profits. Furthermore, agents that are particularly intelligent or well-informed are likely to survive and grow while others fail. Thus in equilibrium a large fraction of observed behavior will be better characterized by the strong assumption that agents know the structure of the economy, than by the equally strong assumption that they form their expectations exclusively by looking backwards.[2]

In a competitive labor market, the argument is somewhat different. Here most of the action comes from the worker's problem of choosing the best model (i.e., firm) in which to participate. Since a searching worker does not typically get to observe a firm very closely before he joins it, it does not seem sensible to assume he has intimate knowledge of its technology or the tastes of its owners. He may, however, have some idea of how it has treated workers in the past. If the worker sees no reason for the firm's technology to suddenly change, and if the worker himself is not very different from other workers the firm has hired in the past, then he may do very well just by assuming that the firm will treat him as it has treated everyone else.

Of course, if he does have more information, then it will be in his interest to use it. It is not the case, however, that such intelligent or well-informed agents will dominate the sample in the way they do in financial markets. In fact, successful agents may leave the sample if they find themselves jobs with some rents and stop searching.[3]

[2]The use of rational expectations is also philosophically appealing in that it is completely free of paternalism. The present paper might seem to be guilty on this account since the inefficiency it identifies is partly due to the inability of workers to predict firms' future behavior by solving their maximization problems for each state of the world. However, this is only paternalistic if we argue that governments can somehow overcome the inefficiency by solving problems that the workers cannot. This is definitely not the approach taken in Section III.

[3]The competitive nature of the labor market is clearly essential to the argument. If workers can band together to threaten the firm, then clearly they can influence its behavior directly. In this model the workers' only options are to "enter" or "exit." The effect of unions is discussed in Section III. In spite of the preceeding paragraphs in the text, it is clear that the question of how best to model behavior cannot be settled on heuristic or even purely theoretical grounds. The transactions costs associated with different institutional structures cannot be compared at a theoretical level with the allocational costs also associated with these structures. Ultimately we can only build models with different behavioral assumptions and compare their implications with the data. This is done (albeit informally) in Sections II and III.

We are ready to begin. The structure of this labor market is not unlike that used in some recent work on employment contingent wage contracts. There are many firms, not necessarily identical, that hire workers to produce output. Firms expect to exist for infinitely many periods, hiring workers at the beginning of each period.[4] At the time of hiring, the state of the world which will prevail for the next period (or any subsequent period) is not known. There are finitely many of these states, indexed by the price of a firm's output P_i, $i=1,\dots,n$. Each price occurs with probability π_i. Workers are identical and production is separable, so the value of each worker's output in state i can be written $P_i X(h_i)$ where h_i is the number of hours he works and X is the amount he produces in physical units; $X(h)$ is concave.

The workers in this model face a short-run mobility cost that prevents them from leaving the firm the moment the state of the world is revealed. For analytical convenience, the length of time they are required to stay is set equal to one period (i.e., the length of time any given state of the world remains in effect). Between periods they may reallocate themselves among firms based on the expected utility each firm offers them. The length of one period (in real time) will have an important effect on the results.

Workers are risk neutral, wishing to maximize a utility function of the form $U(Y,h) = Y - c(h)$, where Y is their income and $c(h)$ is convex.[5] When a worker is hired, the firm may pay him an amount Y_0 as a form of "signing bonus." Once the state of the world is revealed, the firm will pay an additional amount Y_i and tell the worker the number of hours he is expected to work. The worker is assumed to accede to this request, although if he does not like it he may leave at the end of the period. Between periods, workers will therefore migrate to those firms where they perceive $Y_0 + \Sigma\pi_i[Y_i - c(h_i)]$ to be the highest.

Firms' reputations, as discussed earlier, enter into the model in the way workers form their expectations about the expected utility each firm offers. Workers are assumed to possess very good information about the way each firm has treated its workers in the past. They also are aware of the distribution of states of the world, and observe the states when they are revealed.[6] They believe (correctly) that each firm's technology is stable, so that their best predictor of what the firm will do in a given state is simply what the firm did the last time that state occurred. We assume that each firm has existed long enough for the workers to have had at least one observation of its behavior in each state.

Consider first a simplified version of the model where firms offer a noncontingent income level to workers, but retain the right to set hours of work after the state is revealed. Since workers may move freely among firms between periods, in order to hire (and retain) workers, each firm must offer the market an expected utility that at least matches the market level. In any period s this hiring constraint is given by

$$(1) \qquad \overline{U} = \sum_{i=1}^{n} \pi_i\left[Y_s - c\left(h_i^*\right)\right],$$

where h_i^* is the number of hours the firm offered workers the last time that state i occurred. This constraint is not the usual one found in implicit contract models. Note in particular that the firm has only one instrument at the beginning of period s with which

[4]It is important for the results that follow that a firm maximizing the present value of its profits expect to stay in business in future periods. This requires the existence in equilibrium of rents or quasi rents. Since these issues have been treated elsewhere (Klein and Leffler; Shapiro, 1981), they will not be discussed further here.

[5]Since workers are risk neutral, it is the mobility cost that ensures that workers' expectations will be important in the model, and provides a role for reputations, or contracts, to play. Without the mobility cost, the market would become a pure auction market.

[6]This assumption is unrealistic, given the other constraints on the workers' knowledge, but it is most useful as a simplification. Otherwise it would be necessary to model the process by which workers update their beliefs about the distribution of states of the world. The results of this paper would, if anything, become sharper in this circumstance, since the reputational mechanism will be even less efficient and it will be even more difficult for workers to evaluate explicit contracts.

to modify the value of its offer. It is for this reason that a proper dynamic reputational model will not be equivalent to the *ex ante* bargaining setup familiar to contract theory. Contract theory allows firms to adjust the wage and employment components of their "implicit" agreement contemporaneously. A fundamental property of a reputation, however, is that it takes time for it to be built up or run down.

The firm discounts future profits by the factor δ per year. When there are m periods in a year, $m \geq 1$, we may approximate the discount factor per period by $\delta(m)=1-(1-\delta)/m$.[7] The present value of the firm's profits in period t, assuming arbitrarily that state 1 has occurred, is therefore given by

$$(2)\quad V_t = P_1X(h_1) - Y_t + \sum_{s=t+1}^{\infty} \delta(m)^{s-t} \sum_{i=1}^{n} \pi_i [P_iX(h_i) - Y_s].$$

The firm's problem in the current period is to choose Y_t and h_1 to maximize V_t. However, Y_t is predetermined by the constraint (1). In order to attract workers, the firm must pay an amount determined by its past behavior. In the same way, the choice of how many hours to offer is constrained by the effect it will have on future wages. This effect of present behavior on future costs will last until state 1 occurs again. The first-order condition for h_1 is therefore given by

$$(3)\quad P_1X'(h_1) - [\delta(m)\, dY_{t+1}/dh_1]\pi_1 - [\delta(m)\, dY_{t+1}/dh_1 + \delta(m)^2\, dY_{t+2}/dh_1](1-\pi_1)\pi_1 - [\delta(m)\, dY_{t+1}/dh_1 + \delta(m)^2\, dY_{t+2}/dh_1 + \delta(m)^3\, dY_{t+3}/dh_1](1-\pi_1)^2\pi_1 - \ldots = 0$$

Adding contemporaneous terms, this becomes

$$(4)\quad P_1X'(h_1) - \sum_{s=t+1}^{\infty} \delta(m)^{s-t}(1-\pi_1)^{s-t-1}\, dY_s/dh_1 = 0$$

Using (1) this can be written as[8]

$$(5)\quad P_1X'(h_1)[1-\delta(m)(1-\pi_1)] = \pi_1\delta(m)c'(h_1).$$

Since Y_t does not appear in (5), it is clear that the firm's optimal employment strategy in each state will not depend on what has happened in previous periods. Thus if $\bar{U}$ and $\delta(m)$ remain constant (along with tastes and technology), then Y_t and h_i will converge to their steady-state values at their earliest opportunity.

We see from equation (5) that the profit-maximizing strategy of a firm in this reputational market will not in general involve setting employment levels efficiently. This is true even though in the steady state all workers' expectations about employment levels at each firm are correct.[9] The only time when a firm will behave efficiently is

[7]The approximation is easily derived by equating the present value of a constant stream of income from the present to infinity discounted once per period to the present value of the same stream discounted m times per period.

[8]The second term in (4) reflects the concern the firm has for the effect of current policies on future wage costs. In practice, the firm will be learning about these effects just as the workers are learning about the firm's policies. This is a difficult process to model explicitly since we will be concerned mainly with steady-state solutions, and the firm does not learn anything new in the steady state. However, any information the firm does have about the size of these future wage effects will have been generated by the constraint (1). This paper will therefore consider only those solutions in which the firm's conjectures are consistent with (1). One advantage of this approach is that it generates equilibria that are not affected by what the firm might learn from small accidental or experimental deviations from its steady-state strategy.

[9]This implication is common to many reputational models (Shapiro, 1982), but does not seem to be well understood in the labor economics literature. There the argument that correct information in steady state guarantees efficiency seems to be accepted. This is true in particular of the industrial safety literature (Samuel Rea, 1981).

when $\delta(m)=1$. This will occur if $\delta=1$, or as $m \to \infty$ (i.e., as the length of time the workers are attached to the firm becomes very short). Otherwise there will be overemployment.

The intuition behind the role of the discount factor is reasonably straightforward. To the firm, building a reputation is an investment just like any other. The costs of the investment are the present period profits foregone by not following the purely myopic strategy of setting hours such that $P_1 X'(h_1) = 0$. The returns are the lower wage costs achieved in future periods through the effect of current behavior on workers' expectations. The extent to which a firm will invest in its reputation will therefore clearly depend on the interest rate it faces. Of course, this is true of all investments and does not usually indicate the presence of an inefficiency. The difference here is that the cost of developing a reputation, although a private cost to the firm, is not a social cost. There can be no net social costs associated with behaving efficiently for one period. The "cost" of a reputation is a direct transfer to incumbent workers.

The efficiency of the firm's policies also depends on π_1, the probability of the current state occurring again. The intuition is again straightforward. If a particular state is not very probable, then a low level of utility in that state will not reduce expected utility by very much, and therefore not increase workers' asking wages by very much. The private cost to the firm of myopic behavior is therefore lower.

It is interesting to compare this outcome with what could be achieved in a world of explicit contracts. Suppose, for example, that the firm could guarantee that income each period s would be given by $Y_s = Y_0 + c(h_i)$. From (1) it is apparent that the worker will be completely indifferent to the hours offered by the firm in the current period. Future wage costs will be unaffected by the current choice of hours, and (3) becomes

$$P_1 X'(h_1) - c'(h_1) = 0. \tag{6}$$

The firm will therefore set employment levels efficiently in each state.

Several authors have speculated that a firm's concern about its reputation might support a wage function like the one above.[10] If this is the case, then the extension of the results from explicit contract models to the nonunion sector of the labor market might be justified. We can investigate this question by giving the firm in this model the latitude to change the wages of workers in the current period, subject to the effects this may have on future hiring costs. For this purpose, let a worker's income in period s be $Y_s + Y_i + w_i h_i$. The level of nonwage income is Y_i, the marginal wage received in state i is W_i, and Y_s is a noncontingent "signing bonus," as before. The hiring constraint in period s becomes

$$\overline{U} \le Y_s + \sum_{i=1}^{n} \pi_i \left[Y_i^* + w_i^* h_i^* - c(h_i^*) \right]. \tag{7}$$

In the current state, Y_t is predetermined as before, and the firm chooses Y_1, w_1, and h_1 to maximize

$$V_t = P_1 X(h_1) - Y_1 - w_1 h_1 - Y_t + \sum_{s=t+1}^{\infty} \delta(m)^{s-t} \sum_{i=1}^{n} \pi_i \left[P_i X(h_i) - Y_i - w_i h_i - Y_s \right]. \tag{8}$$

Assume there is a lower bound on Y_1 and w_1 (zero is a good candidate) and let the multipliers associated with these lower bounds be λ and γ, respectively. The first-order condition for Y_1, following the methodology of the last example, works out to

$$-1 - \left[\delta(m)\, dY_s / dY_1 \right] / \left[1 - \delta(m)(1-\pi_1) \right] + \lambda = 0, \tag{9}$$

so that (using (7)) we have

$$-(1-\delta(m)) + \lambda \left[1 - \delta(m)(1-\pi_1) \right] = 0, \tag{10}$$

$$(Y_1)\lambda = 0, \tag{11}$$

while for w_1, in the same way, we obtain

$$-(1-\delta(m)h_1 + \gamma \left[1 - \delta(m)(1-\pi_1) \right] = 0, \tag{12}$$

$$(W_1)\gamma = 0. \tag{13}$$

[10]In particular V. V. Chari (1983) and Hart. Both authors are careful to state that their comments are speculative.

These conditions imply that the discretionary part of the current wage will always be set as low as possible. When this lower bound is zero, it is easy to see that the first-order conditions for h_1 will be exactly the same as in the last example. In the steady state, workers will receive a constant salary and will be overemployed just as before.

This result that salaries will be noncontingent in the steady state is interesting because it appears to offer a completely new explanation for the slow adjustment of wages to the state of demand at a firm. It arises because even in a reputational market, a firm must be able to offer its employees *some* guarantees. (If it cannot, a rise in $\bar{U}$ will put the firm out of business, since it will be unable to do *anything* to increase the expected utility of its offer.) The critical assumption which is being made here is that promises to the effect that the firm agrees to pay Y_t if the worker stays for one period are simple enough to be clearly understood and evaluated, while those that outline an employment contingent wage policy are not. As in the early implicit contract models, wages are only constant in the steady state. If the distribution of states of the world were to shift, or if $\bar{U}$ were to change, then the steady-state value of wages would also change. The movement of wages between steady states is discussed further in the next section and in the Appendix.

The failure of a contingent wage policy to promote efficient employment levels is stark evidence of the difference between reputational markets and explicit contract markets. The result is surprising because normally we think of the current wage as an instrument for affecting the firm's choice of current employment levels. However, in this context, the firm already has direct control of employment levels. The choice of w_1 here will affect the choice of h_1, but since hours are set optimally (for the firm) anyway, this further influence from wages has no marginal effect on expected profits. The firm simply sees the lowering of current wages as a way of borrowing money interest free from its workers.

As a final note to this model, and in anticipation of Section III, it is observed that wages seem to change with hours of work at nonunion as well as union firms. Overtime premia are perhaps the best example. It is clear from the model that some form of overtime pay can be useful in that it will induce firms to set the marginal value of the worker's time closer to its marginal cost. Unfortunately, firms in a reputational market are unlikely to provide it. One answer would seem to be laws which require overtime pay. These laws exist, of course, and one wonders why they were perceived to be necessary if firms were already providing it as part of their efficient implicit contracts.

II. A Model with Layoffs and Severance Pay

This section briefly works out a model where the firm controls the number of workers it employs *ex post*, rather than the hours of each worker. This is done simply to generate some more empirical predictions, since the theoretical issues which arise are the same as those in the previous section. The major implications involve the effect of interest rates on employment levels, and the importance of severance pay for laid-off workers. Since examination of the dynamics is reasonably complicated, some of the work is relegated to the Appendix.

Most of the notation and structure of this model carry through from the last section. The firm chooses to hire each period the number of workers that will maximize its expected profits. Suppose this leads to N workers being hired. ($\bar{U}$ will adjust so that all workers will be attached to some firm at the beginning of each period.) After the state is revealed, the firm decides to retain R_i of them. Workers who are laid off in state i receive severance pay s_i.[11] Workers are risk averse, wishing to maximize the utility function $U(Y)$. If employed in state r they receive the wage w_r, and if unemployed, they receive the value of their time at home v plus

[11] The level of severance pay is assumed here to be contingent on the state of the world. It is tempting to analyze the model under the assumption that workers believe the level of severance pay to be noncontingent, in that different numbers of workers will be laid off in different states, but that each will receive the same amount. It turns out, however, that it is not in the firm's *ex post* interest to maintain the level of severance pay constant. These expectations will therefore not be sustained in equilibrium.

any severance pay they may have received. In order to hire workers in state r, the firm must therefore offer an income level w_r which satisfies

$$(14)\quad \overline{U} \leq \sum_{i=1}^{n} \pi_i \big[U(w_r) R_i^*/N + U(s_i^* + v)(N - R_i^*)/N \big].$$

Let us assume $\overline{U} > U(v)$. The firm in period t chooses w_t, s_1, and R_1 to maximize its expected profits, given by

$$(15)\quad V_t = P_1 X(R_1) - w_t R_1 - s_1(N - R_1) + \sum_{r=t+1}^{\infty} \delta(m)^{r-t} \sum_{i=1}^{n} \pi_i \big[P_i X(R_i) - w_r R_i - s_i(N - R_i) \big],$$

where w_t is predetermined, as before. Assume the firm cannot charge workers severance pay when it lays them off, and let the multiplier associated with this constraint be λ. The first-order condition for s_1, following the methodology of the last section, is given by

$$(16)\quad -(N - R_1) - \sum_{r=t+1}^{\infty} \delta(m)^{r-t}(1-\pi_1)^{r-t-1}[dw_r/ds_1] \times \sum_{i=1}^{n} \pi_1 R_1 + \lambda = 0.$$

In the steady state, $w_r = w$ for all r. Thus, using (14) and the lower bound on s_1, we get

$$(17)\quad \pi_1 \delta(m) U'(s_1 + v) - [1 - \delta(m)(1 - \pi_1)] U'(w) + \lambda [1 - \delta(m)(1-\pi_1)] U'(w)/(N - R_1) = 0$$

and

$$(18)\quad (s_1)\lambda = 0.$$

Consider first the case where $\delta(m) = 1$. If we assume there is no constraint on severance pay, we see from (17) that $s_1 = w - v$, which satisfies the constraint. If $\delta(m)$ falls with $\lambda = 0$, $U'(s_1 + v)$ must rise relative to $U'(w)$. Thus in general, reputational markets may provide risk-averse workers with severance pay, but not enough to insure them completely from the risk of layoff. If workers are risk neutral, note that severance pay will never be provided.

One of the problems with the early work on implicit contracts was its prediction that laid-off workers should always be provided with enough severance pay to completely insure them from the risk of layoff. In practice, severance pay is not nearly so widespread, especially in the nonunion sector. Once reputations are explicitly modelled, however, we see that the problem disappears. Severance pay will only arise in reputational markets if interest rates are low enough, and even then it will be underprovided. Note that severance pay does not enter this model as an instrument to control the firm's layoff decision. The firm controls this directly and does not need any help. Rather, severance pay exists purely as a form of insurance for laid-off workers. It is underprovided when interest rates are positive because, as in the last section, firms see reductions in severance pay as a way to borrow money interest free from their workers.

It is interesting that some firms offer severance pay to workers who quit, if notice is given (Bureau of National Affairs, 1978). This evidence is consistent with an insurance-based model of severance pay, but not with a model that argues it is an instrument for reducing layoffs.

Examination of the retention decision again follows the methodology of the last section. The first-order condition for R_1 in a state where $R_1 < N$ is given by[12]

$$(19)\quad P_1 X'(R_1) - w_t + s_1 - \sum_{r=t+1}^{\infty} \delta(m)^{r-t}(1-\pi_1)^{r-t-1} \times [dw_t/dR_1] \sum_{i=1}^{n} \pi_i R_i = 0.$$

[12] This constraint is not being explicitly dealt with here for simplicity. As in Azariadis-type contract models, it will typically be binding in good periods. If v is

Using (14) in the steady state, this works out to

$$(20)\quad [P_1X'(R_1)-w+s_1][1-\delta(m)(1-\pi_1)]U'(w)+\pi_1\delta(m)[U(w)-U(s_1+v)]=0.$$

In the Appendix, the following results are proven. When severance pay is provided, the dynamic system defined by (14), (17), and (20) is stable in a region around $\delta(m)=1$ (which is precisely the region where severance pay is likely to be provided). Increases in interest rates will reduce employment in this region minus the single point $\delta(m)=1$. When the severance pay constraint is binding, stability is harder to guarantee. However, when the system is stable, increases in interest rates will again reduce employment in the steady state. During transition periods, the model may exhibit cobweb-type cycles.

One of the interesting (although perhaps not very robust) properties of the model concerns the behavior of wages during transition periods. This property follows directly from the form of the expected utility constraint (14). If an exogenous variable which affects the steady-state demand at the firm (such as interest rates or the overall distribution of states of the world) changes, then the steady-state levels of wages and employment also change. However, the first movement is always in employment levels, with wages remaining "sticky." This is simply because workers' expectations about employment levels only change after the firm's behavior changes. On the other hand, a change in $\overline{U}$ (supply) will always affect wage levels first. However it must be noted that if workers can observe the change directly and predict its effect, then movements should be directly to the new steady state. As well, the expectational rule used by workers is not a very good one outside of the steady state.[13]

Unlike the unemployment which would arise in an Azariadis-Baily world, the joblessness in this reputational world is both involuntary and inefficient *ex post*. Furthermore, it increases as interest rates increase.[14] Of course, it is dangerous to generalize from a partial equilibrium model, but the intuition behind this result seems quite robust. There is a great deal of evidence from the labor market to support the existence of labor hoarding during recessions. By definition, if a firm is hoarding labor, it is retaining more workers than it needs to maximize short-run profits because it sees the prospect of a longer-run gain. In this model, that gain involves reputation, while in the standard model it involves the return on the specific capital imbodied in those extra workers. In either case, as the opportunity cost of the funds used to pay these workers increases, we can only expect more of them to be let go.

The relationship between interest rates and employment has been the subject of several recent empirical investigations, including one by Orley Ashenfelter and David Card (1982). In a vector autoregressive study of the time-series data, these authors find a marked similarity between the paths of employment and nominal interest rates, with strong evidence of correlation even in the very short run. The

low enough relative to the lowest P_i, and interest rates are low, then the constraint will be binding in all periods. On the other hand it is possible for there to be layoffs even when v is zero, if interest rates are high enough.

[13]Benjamin Eden (1983) discusses this point.

[14]The model of this section predicts underemployment while that of the previous section predicts overemployment. The reason is that in the first model, the firm is committed to the payment of a fixed salary regardless of the hours worked by its employees. In the second, the firm can save this amount by setting hours to zero. This paper has little to say on the question of why white-collar workers work on salary, enjoy stable employment (under normal circumstances), and have variable amounts of work to do, while blue-collar workers are regularly subject to temporary layoff. James Markusen (1979) provides an explanation in a contracting framework. However, given the existence of these two modes of employment, the models make the rather precise prediction that an increase in interest rates, holding the state of demand constant, will lead to more work being asked of white-collar employees while at the same time more blue-collar workers are laid off.

direction of the relationship is consistent with the present model (and the opposite of that predicted by the simplest Lucas-Rapping (1969) supply function models). Furthermore, the short-run nature of the relationship seems to indicate the effect is not operating through the rental price of capital. Of course, this model predicts a relationship between employment and *ex ante* real interest rates, but innovations in nominal rates may be a better proxy for this than innovations in *ex post* real rates (which seem unrelated to employment in the Ashenfelter-Card study). Further investigation, including a proper general equilibrium analysis, is needed, but it appears there may be a strong direct link between money and employment which has hitherto escaped attention.

III. Unions and Government Policy

The labor markets described in this paper seem to bear a stronger resemblance to those one might find in a novel by Charles Dickens than to those which appear in any modern industrial relations text. Can the predictions of this paper be reconciled with the plethora of institutions which we observe in modern labor markets? The following interpretation suggests itself.

Pure reputational markets do not work very efficiently. The reason is that firms are not rewarded soon enough. They must wait until workers' expectations have changed before they can expect to see a wage reduction in return for their good behavior. One way a firm might hope to get around this problem would be to offer each worker, as he is hired, an explicit contract which commits the firm to a certain pattern of future behavior. In this way it might be able to get a wage reduction right away. Contracts of this sort, however, just do not seem to be a part of the modern labor market. The suggestion offered here is that in the complex real world, they would either be too difficult to police *ex post* (as with state contingent employment contracts), or too difficult to interpret or evaluate *ex ante* (as with employment contingent wage contracts).

Workers since the industrial revolution have been attempting to increase their welfare through unionization. Once a union has established itself, and its agents are bargaining at regular intervals with the firm, the environment changes to more closely resemble that of the explicit contract models. Binding commitments are possible, and over time the information and experience necessary to evaluate them may become available. It is not completely unreasonable that as bargaining continues, the institutions agreed upon at union firms will evolve toward efficiency. It does not follow, of course, that firms should be inviting their workers to unionize, since in the process they lose the advantages of dealing with a competitive labor market. It does follow, however, that union workers may be able to command higher wages than their nonunion counterparts even in highly competitive industries.

The existence of institutions like overtime pay in the nonunion sector, according to this approach, must be blamed on government. Here the efficiency argument becomes more problematic since it is unlikely that government could ever collect the detailed, firm-specific information necessary to design an efficient institution. However, the firms in this sector, even though they may have better information than a central planner, do not have the incentive to use it. As well, extra information is available in the form of the outcomes of previous collective bargaining agreements. Government may be able to improve on things simply by imposing on the nonunion sector some of the institutions spawned in the union sector. Once the government has entered the market in this way, the incentive to unionize in highly competitive industries should become weaker.

Of course there are still many problems, and this section should not be read as a justification of current levels of government intervention in the labor market. It may be, however, that the present high level of intervention can be partly *explained* by lack of confidence on the part of legislators in the reputational mechanism. Governments try to use legislation to increase safety levels in mines, for example, since they do not believe that a mine with a bad safety record will be so worried about having to pay high wages in the future that it will institute the required

programs itself. This paper indicates that their reasoning may be correct, especially since information about changes in the probability of an accident at a firm may be quite slow to disseminate. However, it is worth emphasizing again that the models of this paper suggest only that there may be a role for a benign, well informed, and efficiently run central planning authority. They say nothing about the cost effectiveness of current programs.

IV. Conclusions

This paper has examined two simple models of the labor market. The models build on the realistic behavioral assumption that firms do not enter into explicit contracts with their workers, but that they do develop reputations for their policies. The models shed light on many aspects of the modern labor market, including the insensitivity of wages to the state of demand at a firm, the generally low levels of severance pay which are received by laid-off workers, the growth of unions, and the role of government policy. They also highlight one possible mechanism for the transmission of monetary stimuli to the real sector of the economy. Finally, they show that the old maxim that there is no reason for parties to bargain themselves to a point not on the contract curve simply does not apply when "contracts" are implicit. Recognition of this may open up a whole new area for policy-oriented research—an area which is completely hidden by current models of the labor market.

Appendix

This Appendix examines the dynamics and some comparative statics of the model of Section II. The structure of the system is given by (14), (17), and (20) with the constraint (18). The first three of these are reproduced below.

$$\text{(A1)}\quad \bar{U} = \sum_{i=1}^{n} \pi_i \left[U(w_{t+1}) R_{1,t}/N + U(s_{1,t}+v)(N-R_{1,t})/N \right],$$

$$\text{(A2)}\quad \pi_1 \delta(m) U'(s_{1,t}+v) - [1-\delta(m)(1-\pi_1)]\, U'(w_t) + \lambda = 0,$$

$$\text{(A3)}\quad [P_1 X'(R_{1,t}) - w_t + s_{1,t}] \times [1-\delta(m)(1-\pi_1)]\, U'(w_t) + \pi_1 \delta(m)[U(w_t) - U(s_{1,t}+v)] = 0.$$

Consider first the case where the severance pay constraint is not binding. Totally differentiating the system at a steady-state solution we get the following useful matrix equation:

$$\text{(A4)}\quad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \begin{vmatrix} dw \\ ds_1 \\ dR_1 \end{vmatrix} = \begin{vmatrix} b_1 \\ b_2 \\ b_2 \end{vmatrix} d\delta(m),$$

where

$$\begin{aligned}
a_{11} &= \Sigma \pi_i R_i U'(w) \\
a_{12} &= \pi_1 U'(s_1+v)(N-R_1) \\
a_{13} &= \pi_1 [U(w) - U(s_1+v)] \\
a_{21} &= -[1-\delta(m)(1-\pi_1)] U''(w) \\
a_{22} &= \pi_1 \delta(m) U''(s_1+v) \\
a_{23} &= 0 \\
a_{31} &= [P_1 X'(R_1) - w + s_1][1-\delta(m)(1-\pi_1)] \\
&\quad U''(w) - (1-\delta(m)) U'(w) \\
a_{32} &= 0 \\
a_{33} &= P_1 X''(R_1)[1-\delta(m)(1-\pi_1)] U'(w) \\
b_1 &= 0 \\
b_2 &= -\{\pi_1 U'(s_1+v) + (1-\pi_1) U'(w)\} \\
b_3 &= (1-\pi_1)[P_1 X'(R_1) - w + s_1] U'(w) \\
&\quad - \pi_1 [u(w) - U(s_1+v)].
\end{aligned}$$

Linearizing the original system around a steady state value, we get

$$\text{(A5)}\quad w_{t+1} = -[a_{12}/a_{11}] s_{1,t} - [a_{13}/a_{11}] R_{1,t} + A$$

$$\text{(A6)}\quad s_{1,t} = -[a_{21}/a_{22}] w_t + B$$

$$\text{(A7)}\quad R_{1,t} = -[a_{31}/a_{33}] w_t + C,$$

where A, B, and C are constants. We can

now easily express w_{t+1} as a function of w_t.

$$\text{(A8)} \quad w_{t+1} = \{[a_{12}a_{21}]/[a_{11}a_{22}] + [a_{13}a_{31}]/[a_{11}a_{33}]\}w_t + A + B + C.$$

We cannot guarantee in general that the term in braces in (A8) will be less than one in absolute value. However, as $\delta(m)$ approaches one, the first-order conditions tell us that $s_1 + v - w$ approaches zero. Thus a_{13} and the second term in the brackets will disappear as interest rates approach zero. The first is negative, but will be greater than minus one under the usual assumption that absolute risk aversion is constant or declining. The system should therefore be stable in a region around $\delta(m) = 1$.

For the comparative static results between steady states, note that the sign pattern of (A4) is given by

$$\begin{vmatrix} (+) & (+) & (+) \\ (+) & (-) & 0 \\ (?) & 0 & (-) \end{vmatrix} \begin{vmatrix} dw \\ ds_1 \\ dR_1 \end{vmatrix} = \begin{vmatrix} 0 \\ (-) \\ (-) \end{vmatrix}$$

The determinant is therefore likely to be positive, but this cannot be guaranteed due to the indeterminancy of a_{31}. However, as interest rates approach zero, a_{13} also goes to zero. The determinant is therefore positive in this case. Also, b_3 goes to zero, however, so by Cramer's Rule, interest rates have no effect on employment at this point. However, if interest rates are positive and the determinant remains positive, it is easy to see that $dR_1/d\delta(m) > 0$ regardless of the sign of a_{31}. Increases in interest rates (i.e., a fall in $\delta(m)$) will reduce steady-state employment.

Consider now the case where the severance pay constraint is binding. The system is defined by (A1) and (A3) with $s_1 = 0$. We have

$$\text{(A9)} \quad \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} \begin{vmatrix} dw \\ dR_1 \end{vmatrix} = \begin{vmatrix} 0 \\ b_3 \end{vmatrix} d\delta(m),$$

$$\text{(A10)} \quad w_{t+1} = [a_{13}a_{31}/a_{11}a_{33}]w_t + D.$$

Again, we cannot be sure that the term in brackets is less than one in absolute value. The ambiguity in a_{13} is due to the fact that if workers are very risk averse, a rise in the wage rate may increase the risk of the firm's offer so much that the firm increases employment in the current period in order to mitigate this effect. Under risk neutrality, a_{31} is therefore negative. As risk aversion increases, a_{31} moves toward zero and the system is sure to be stable. As a_{31} continues to grow and becomes positive, the model will exhibit cycles, and finally explosive cycles.

The sign pattern for the matrix in (A9) is

$$\begin{vmatrix} (+) & (+) \\ (?) & (-) \end{vmatrix} \begin{vmatrix} dw \\ dR_1 \end{vmatrix} = \begin{vmatrix} 0 \\ (-) \end{vmatrix} d\delta(m).$$

The determinant is therefore likely to be negative, but again it cannot be signed. However, so long as the system is stable the determinant will be negative. A simple application of Cramer's Rule shows once again that $dR_1/d\delta(m) > 0$ regardless of the sign of a_{31}, so increases in interest rates again reduce employment levels in the steady state.

It is also interesting to trace through some of the short-run effects of changes in the exogenous variables. Although a change in interest rates will eventually affect wages, the initial effect is felt only on employment levels and severance pay. Conversely, a change in supply conditions will lead to an immediate change in wages, and only later will it affect employment levels.

REFERENCES

Ashenfelter, Orley and Card, David, "Time Series Representations of Economic Variables and Alternative Models of the Labor Market," *Review of Economic Studies*, December 1982, *49*, 761–83.

Azariadis, Costas, "Implicit Contracts and Underemployment Equilibria," *Journal of Political Economy*, December 1975, *83*, 1183–202.

______, "Employment with Asymmetric Information," *Quarterly Journal of Economics*, Supplement, 1983, *98*, 157–73.

Baily, Martin Neil, "Wages and Employment with Uncertain Demand," *Review of Economic Studies*, January 1974, *41*, 37–50.

Barro, Robert and Gordon, David, "Rules, Dis-

cretion, and Reputation in a Model of Monetary Policy," *Journal of Monetary Economics*, July 1983, *12*, 101–21.

Chari, V. V., "Involuntary Unemployment and Implicit Contracts," *Quarterly Journal of Economics*, Supplement, 1983, *98*, 107–23.

Cooper, Russell, "A Note on Overemployment/Underemployment in Labor Contracts Under Asymmetric Information," *Economics Letters*, January 1983, *12*, 81–89.

Eden, Benjamin, "Reputation and Aggregate Layoffs," mimeo., University of Iowa, December 1983.

Gordon, Donald F., "A Neo-Classical Theory of Keynesian Unemployment," *Economic Inquiry*, December 1974, *12*, 431–59.

Grossman, Sanford and Hart, Oliver, "Implicit Contracts, Moral Hazard, and Unemployment," *American Economic Review Proceedings*, May 1981, *71*, 301–07.

Hall, Robert E. and Lilien, David, "Efficient Wage Bargains Under Uncertain Supply and Demand," *American Economic Review*, December 1979, *69*, 868–79.

Hart, Oliver, "Optimal Labor Contracts Under Asymmetric Information: An Introduction," *Review of Economic Studies*, January 1983, *50*, 3–36.

Holmstrom, Bengt, "Equilibrium Long Term Labour Contracts," *Quarterly Journal of Economics*, Supplement, 1983, *98*, 23–54.

______, "Contractual Models of the Labor Market," *American Economic Review Proceedings*, May 1981, *71*, 308–13.

Klein, Benjamin and Leffler, Keith, "The Role of Market Forces in Assuring Market Performance," *Journal of Political Economy*, October 1981, *89*, 615–41.

Lazear, Edward, "Agency, Earnings Profiles, Productivity and Hours Restrictions," *American Economic Review*, December 1981, *71*, 606–20.

Lucas, Robert and Rapping, Leonard, "Real Wages, Employment, and Inflation," *Journal of Political Economy*, September/October 1969, *77*, 721–54.

Markusen, James, "Personal and Job Characteristics as Determinants of Employee-Firm Contract Structure," *Quarterly Journal of Economics*, April 1979, *93*, 255–80.

Radner, Roy, "Monitoring Cooperative Agreements in a Repeated Principal Agent Relationship," *Econometrica*, December 1981, *49*, 1127–48.

Rea, Samuel, "Workman's Compensation and Occupational Safety under Imperfect Information," *American Economic Review*, March 1981, *71*, 80–93.

Shapiro, Carl, "Premiums for High Quality Products as Rents for Reputation," Discussion Papers in Economics No. 6, Woodrow Wilson School, Princeton University, 1981.

______, "Consumer Information, Product Quality, and Seller Reputation," *Bell Journal of Economics*, Fall 1982, *13*, 20–35.

Strand, Jon, "The Structure and Efficiency of Reputation Labor Contracts," Working Paper No 46, Stanford University Workshop on the Economics of Factor Markets, 1983.

Townsend, Robert, "Optimal Multiperiod Contracts and the Gain from Enduring Relationships Under Private Information," *Journal of Political Economy*, December 1982, *90*, 1166–85.

Bureau of National Affairs, "Separation Procedures and Severance Benefits," Personnel Policies Forum No. 121, April 1978.

LDC Foreign Borrowing and Default Risk: An Empirical Investigation, 1976–80

By SEBASTIAN EDWARDS*

The recent foreign debt crisis faced by some less developed countries (*LDCs*) (Mexico, Brazil, Argentina, among others) has generated concern among economists, bankers, and politicians. In particular, the ability of the international banks to distinguish between "good" and "bad" risks has been questioned. It has even been suggested that the inability to restrict credit to countries with low "credit worthiness" has resulted in the overextension of some major banks and that, as a consequence, this has increased the probability of a global international financial collapse.[1]

The purpose of this paper is to investigate to what extent the international financial community has taken into account the risk characteristics of *LDCs* when granting loans. Specifically, this study analyzes the determinants of the spread between the interest rate charged to a particular country and the London Interbank Borrowing Rate (*LIBOR*). If the financial community distinguishes between countries with different probabilities of default, these perceptions will be reflected in the spreads over *LIBOR*, with riskier countries (i.e., countries with a higher probability of default) being charged a higher risk premium or spread. However, when the perceived probability of default exceeds a given level, a "credit-ceiling" will be reached, and that particular country will be completely excluded from the credit market (Jonathan Eaton and Mark Gersovitz, 1980, 1981a, b; Jeffrey Sachs and Daniel Cohen, 1982; Sachs, 1983). This paper also tries to determine if the international financial community anticipated, as late as 1980, the international debt crisis of 1982–83. This is done by computing the implicit subjective probabilities of default from the econometric analysis of the spreads over *LIBOR*.

The empirical analysis of the determinants of the default risk premium is important for several reasons. First, an understanding of the factors that influence lending behavior is useful for borrowing countries. With this knowledge *LDCs* can take positive steps towards managing their economies in a way such that the perceived default risk is kept at a level compatible with what lenders think is prudent. Second, additional information on how the market assesses default risk will be helpful for determining the probability that the present repayment difficulties faced by some *LDCs* can be transformed into a major global crisis. And third, empirical information on the relationship between the level of the foreign debt and its cost is useful for the analysis of optimal borrowing strategies and of the social rate of discount in an open economy (see Arnold Harberger, 1983.)

*Department of Economics, University of California, Los Angeles, CA 90024, and National Bureau of Economic Research, Cambridge, MA 02138. Financial support from UCLA's Council on International and Comparative Studies is gratefully acknowledged. I have benefited from helpful discussions with John Bilson and Carlos Rodriguez, and from comments by two anonymous referees, Michael Darby, Susan Woodward, and the participants in seminars at the University of Chicago, UCLA, and the World Bank. Steve Feinstein and Evan Tanner provided able research assistance. The research reported here is part of the NBER's Research Studies in International Studies. Any opinions expressed here are my own and not those of the NBER.

[1]See, for example, the *Wall Street Journal* editorial, March 9, 1983. On the recent international debt crisis, see the comprehensive discussion in William Cline (1983). In particular, see his ch. 5 for a discussion on bank's responsibility in the present crisis. See also *Time* magazine (January 10, 1983), *The Economist*, (March 5, 1983), Martin Feldstein (1983), David Folkerts-Landau (1982). The indebtedness situation is particularly critical regarding Latin American debtors. For example, U.S. private banks have "extended credit of more than U.S. $50 billion to Mexico, Brazil and Argentina, an amount that exceeds 80 percent of the banks equity" (Feldstein, p. 2). The extent of the indebtedness crisis is reflected by the fact that in 1982, twenty countries undertook debt renegotiations, while in the second half of the 1970's, an average of only four countries per year renegotiated their debts.

A number of papers have recently analyzed the theoretical determinants of default country risk.[2] Recent work has focused on a number of aspects of the problem. First, the existence of credit ceilings, above which countries cannot borrow, has been explicitly introduced into the analysis (Eaton and Gersovitz, 1980, 1981a, b; Sachs and Cohen, 1982; Sachs, 1983; David Folkerts-Landau, 1982.) Second, variables other than the level of foreign debt, like international reserves, the propensity to invest and the current account deficit have been explicitly considered as affecting the default risk premium (Gershon Feder and Richard Just, 1979; Eaton and Gersovitz, 1980, 1981a; Sachs and Cohen, 1982; Sachs, 1983; myself, 1983). Recent theoretical analyses have also made a distinction between bond and bank foreign financing, and have explicitly introduced the possibility of rescheduling debt payments (Sachs and Cohen, 1982; Sachs, 1982). Finally, it has beeen argued that if borrowers and lenders have different perceptions with respect to the probability of default, the analysis of optimal borrowing strategies would be substantially affected (Harberger, 1976, 1980).

The empirical work on the subject has investigated several aspects of the problem, including the probability of a country rescheduling its payments (Charles Frank and William Cline, 1971; Nicholas Sargen, 1977), and the probability that a particular *LDC* borrower has reached its credit ceiling (Eaton and Gersovitz, 1980, 1981a, b). Generally, those studies that have analyzed lending behavior in international financial markets have found that lenders tend to take into account the riskiness of borrowers in making their lending decisions (Frank and Cline, 1971; Feder and Just, 1977a, b; Feder and Knud Ross, 1982; Sachs, 1981). Some studies, however (Feder and Just, 1977b; Sachs, 1981), have only found a weak and insignificant relationship between the spread and the debt-output ratio. In their paper Feder and Ross used data from the *Institutional Investor* creditworthiness ranking to show that lenders risk perceptions are systematically reflected in the spreads charged in Euromarkets.

The analysis presented in the present paper uses data for several years (1976–80) to investigate the determinants of the subjective probability of default.[3] The sample considered in this paper only includes loans denominated in Eurodollars, thus avoiding the problem of different currency composition of loans, mentioned by Donogh McDonald (p. 630). Also this paper only includes public and publicly guaranteed loans, thus restricting the analysis to the determinants of country risk, as distinct from financial risk. Finally, the present study has considered a larger set of possible determinants of the probability of default than previous work.

I. Empirical Analysis

Assume that, as postulated by Feder and Just (1977b), Eaton and Gersovitz (1980), and Sachs (1981) among others, the spread over *LIBOR* charged on Eurodollar loans reflects the probability of default of a particular country. Then, observed data on the spread can be used to formally analyze the way in which variables like the debt-output ratio, the propensity to invest, the level of international reserves, and others affect the level of this perceived probability. However, before empirically analyzing the determinants of the spread, three important questions should be addressed: What is the exact form of the relationship between the spread and the probability of default? What is the functional form of the probability of default? What are the determinants of this probability?

With respect to the first question—the relationship between the spread and the probability of default—in this paper I assume that the spread can be written in the following form:

$$(1) \qquad s = [p/(1-p)]\gamma,$$

[2] See Donogh McDonald (1982) for an excellent and exhaustive survey on the subject.

[3] Most previous work has used cross-section data for a particular year or quarter. Feder and Just (1977b) used loans data for 8 quarters during 1973–74. The present paper uses the most recent data available, since in 1981 the World Bank stopped the publication of *Borrowing in International Capital Markets* from where the data on the spreads was obtained.

where s is the spread over *LIBOR* charged on a particular loan, p is the (subjective) default probability during the life of the loan, and γ is a variable that captures other elements affecting s. There are several ways to justify the choice of equation (1). For example, this expression can be directly derived from the assumption of risk neutral banks and perfect competition.[4]

Regarding the functional form of p, I follow the standard convention and assume that p has a logistic form:

$$p = \frac{\exp \Sigma \alpha_i y_i}{1 + \exp \Sigma \alpha_i y_i} \tag{2}$$

where the y_is refer to the determinants of the perceived probability of default (i.e., the debt-service ratio, the propensity to invest) and the α_is are the corresponding coefficients. Combining equations (1) and (2) the log of the spread can be written in the following form:

$$\log s = \alpha_0 + \Sigma \alpha_i y_i + \log \gamma. \tag{3}$$

In this section the results obtained from the estimation of an equation of the type of (3) using data on 727 public and publicly guaranteed loans granted to 19 *LDC*s during 1976–80 are reported.[5] The spread variable for each country in a particular year was constructed as a weighted average of spreads actually charged for the individual public and publicly guaranteed loans granted to that particular country, where the weights were given by the value of each loan.[6] The basic data were obtained from various issues of the World Bank's *Borrowing in International Capital Markets*, and are presented in my earlier paper.

Several variables were considered as possible determinants of the spread, including those suggested by a number of models that have recently appeared in the literature. (See Feder and Just, 1977b; Sachs and Cohen, 1982; Sachs, 1983; Eaton and Gersovitz, 1980; my earlier paper) Specifically, the following variables—some of which have also been included in previous empirical work on the subject—were considered as possibly affecting s in the empirical analysis.

1) The debt-output ratio. As has been argued by Sachs and Cohen (1982) and others, it is expected that this variable will have a positive coefficient in the regression analysis. This variable can be considered to be an indicator of the degree of solvency of a particular country. The data on this variable refers to public and publicly guaranteed debt and was obtained from the World Bank *World Debt Tables*.

2) The ratio of debt service to exports. This indicator measures possible liquidity (as opposite to solvency) problems faced by a particular country. It is expected that its coefficient will be positive. Data on this ratio was obtained from the *World Debt Tables*.

3) Ratio of international reserves to *GNP*. This indicator measures the level of international liquidity held by a country and as suggested in my earlier paper, it is expected that its coefficient will be negative. This variable was constructed from data obtained from the IMF's *International Financial Statistics*.

[4] This can be illustrated by assuming the simple case of a one-period loan. Consider that the risk-free interest rate is given by $LIBOR(i^*)$, and that the interest rate charged to a particular country (i) is equal to *LIBOR* plus the spread (s). Assume also that in case of default the principal and interest are completely lost (this is a nonessential assumption). Then, the following equilibrium condition will hold: $(1-p)(1+i) = (1+i^*)$. From here it follows that $s = [p/(1-p)]\gamma$, for $\gamma = (1+i^*)$. Alternatively, more complicated models, like the one in Feder and Just (1977b), can be used to develop expressions similar to (1).

[5] Countries included in the empirical analysis are Argentina, Brazil, Colombia, Ecuador, Greece, Indonesia, Ivory Coast, Korea, Malaysia, Mexico, Morocco, Panama, Philippines, Portugal, Spain, Thailand, Uruguay, Venezuela, and Yugoslavia.

[6] It is important to notice that there are some problems with the quality of this data. The most serious problem is that the interest rate is not the only component of a loan's cost. In particular in the present study —as in those by Feder and Just (1977b) and Sachs (1981), for example—it has not been possible (due to lack of information) to incorporate data on fees and commissions. It should be noted, however, that these fees are typically low compared to the interest rate component of the costs (see, for example, the discussion in Cline, pp. 82–83).

4) Loan duration. This variable is measured in years, and measures the (weighted) average maturity of loans granted to a particular country. As has been shown by Feder and Ross (1982) its a priori sign in the regression analysis is ambiguous. The weighted average was constructed from data reported in *Borrowing in International Capital Markets*.

5) Loan volume. This variable shows the average value of each loan, and was obtained from *Borrowing in International Capital Markets*. Also, a priori, its sign is ambiguous.

6) Propensity to invest. This variable was constructed as the ratio of gross domestic investment of *GDP*, and will tend to capture the country's prospectives for future growth. As is shown in Sachs and Cohen (1982), and in my earlier paper, it should be negatively related to the spread. This indicator was obtained from data reported in the *World Tables* and in *World Development Report* (various issues).

7) Ratio of the current account to *GDP*. Sachs (1981) has argued that this variable will be negatively related to the probability of default. The data on this variable was obtained from various issues of the *World Tables* and *World Development Report*.

8) Average propensity to import. This indicator was constructed as the ratio of imports to *GNP*, and measures the degree of openness of a country. To the extent that a more open country is more vulnerable to foreign shocks (Jacob Frenkel, 1983), it is expected that it will be positively related to the probability of default. This variable was constructed from data obtained from the *International Financial Statistics*.

9) Growth of per capita *GDP*. It has been argued that a higher rate of growth of output will result in a lower probability of default (see Feder and Just, 1977b). Data on this indicator was obtained from *World Tables* and the *World Development Report*.

10) *GNP* per capita. This variable measures the relative economic size of a country. Some authors have argued (i.e., Feder and Just, 1977b) that this variable should have a negative coefficient in equation (3). This variable was obtained from various issues of the World Bank's *World Tables* and *World Development Report*.

11) Rate of inflation. It is possible to argue that, with other things given, a higher rate of inflation indicates a larger probability of a balance of payments crisis, and consequently a higher probability of default (McDonald). This variable was taken from the *International Financial Statistics*.

12) Variability of International Reserves. According to the literature on the demand for international reserves (for example, Frenkel) the more variable are the flows of foreign funds faced by a country, the higher the probability of a balance of payments crisis. Consequently, it is expected that the coefficient of this variable will be positive. This indicator was constructed from data obtained from the *International Financial Statistics*.

13) Rate of Devaluation. This variable summarizes the exchange rate policy followed by a particular country. For a given level of the other variables (in particular inflation and reserves), a higher rate of devaluation will tend to indicate a higher willingness to use exchange rate adjustments to avoid balance of payments crises.[7] It is expected that it will have a negative coefficient in the regression analysis. The data was obtained from *International Financial Statistics*.

14) Government expenditure over *GNP*. It has been suggested that the larger the size of the government sector in a developing country, the higher the probability of a balance of payments crisis.[8] It is then expected that the coefficient of this variable will be positive. The data was obtained from *World Tables* and *World Development Report*.

Other variables were also considered as possible determinants of the probability of

[7]Cline (ch. 1) has indicated that the mismanagement of the exchange rate policy is one of the main causes of the international debt crisis. Deviations of Purchasing Power Parity were included instead of the rate of devaluation as an indicator of exchange rate policy. The results obtained, however, were not different.

[8]See, for example, the discussion in Harberger's and my article (1982). Cline argues that the increase in the Mexican fiscal deficit was one of the main causes for the 1982 debt crisis in that country.

default (oil exporters dummy variables, for example). However, due to space considerations, and since their inclusion did not affect the results in any significant way, the estimates obtained when they were included are not reported here.

Equation (3) was estimated using pooled cross-section time-series data for 19 countries during five years (1976–80). For estimation purposes it was assumed that $\log \gamma_{nt}$ was equal to a constant k plus a random element $u_{nt}(\log \gamma_{nt} = k + u_{nt})$. Following the usual convention in pooled time-series cross-section estimation, it was assumed that this random term u_{nt} was formed of a country-specific random error v_n, with zero mean and variance σ_v^2; a time-specific random element w_t, with zero mean and variance σ_w^2; and an independently distributed random term ε_{nt}, with zero mean and variance σ_ε^2 (see Theodore Anderson and Cheng Hsiao, 1981). Then the equation to be estimated can be written as

(4)

$$\log s_{nt} = \alpha_0 + k + \sum_{i=1}^{m} \alpha_i y_{int} + v_n + w_t + \varepsilon_{nt},$$

where

$$E(v_n^2) = \sigma_v^2; \quad E(w_t^2) = \sigma_w^2; \quad E(\varepsilon_{nt}^2) = \sigma_\varepsilon^2,$$

and $E(v_n w_t) = E(v_n \varepsilon_{nt}) = E(w_t \varepsilon_{nt}) = 0$

$$E(v_n v_m) = 0 \quad \text{for } n \neq m$$

$$E(w_t w_s) = 0 \quad \text{for } t \neq s$$

$$E(\varepsilon_{nt}\varepsilon_{ns}) = E(\varepsilon_{mt}\varepsilon_{nt}) = E(\varepsilon_{nt}\varepsilon_{ms}) = 0.$$

Expression (4) is a typical random-effect error components equation. The results presented in this paper were obtained using the technique suggested by Wayne Fuller and George Batesse (1974). In the estimation $(\alpha_0 + k)$ was combined into a constant β_0. A possible problem with the estimation of (4) is that, to the extent that banks determine the spread and loan duration at the same time, use of Fuller-Batesse's technique would be subject to a simultaneity bias. However, following Feder and Ross, and David Beim (1977) it was assumed that the duration of the loan is determined by banks *prior* to the determination of the spread. This indeed appears to be the case in the Eurocurrency credit markets (see *Euromoney*, September 1978).

Table 1 contains the results obtained from the estimation of equation (1) using Fuller-Batesse's technique. These results are quite satisfactory, both from the point of view of the mean square errors of the regressions, and from the perspective of the signs and level of significance of the coefficients. Broadly speaking, the empirical evidence shows that international lending behavior to *LDC*s tends to take into account some of the economic characteristics of the specific borrowing countries. As may be seen, in all regressions the debt-output ratio is significantly positive, and smaller than one. This result suggests that a higher level of indebtedness will be associated with a higher probability of default and thus, a higher spread over *LIBOR*. This contrasts with results previously obtained by Feder and Just (1977b), using data for eight quarters in 1973–74, and by Sachs (1981) where the coefficient of the debt-output ratio was always insignificant and very low. With respect to the debt-service ratio—that measures potential liquidity problems—its coefficients are also positive, as expected, and in most cases significant either at the 5 or 10 percent level.

The coefficient of the reserves to *GNP* ratio is, as expected, consistently negative, and in most equations it is significant at the 5 percent level. Also, the estimated values of these coefficients are high, indicating that the behavior of the reserves ratio has played an important role in the determination of the perceived probability of default. The main importance of this result is that from a policy point of view, countries that want to reduce the probability of being excluded from the international financial market, due to an increase in the perceived probability of default, should be particularly careful in managing their international reserves. Also, these results suggest that the analysis of the demand for international reserves should incorporate the level of foreign indebtedness as an ad-

TABLE 1—ESTIMATION OF EQUATION (4): FULLER-BATESSE PROCEDURE, 1976–80

Independent Variable	Equations					
	(4.1)	(4.2)	(4.3)	(4.4)	(4.5)	(4.6)
Constant	0.329	0.141	0.305	0.285	0.345	0.314
	(1.422)	(0.726)	(1.216)	(1.225)	(1.062)	(1.424)
Debt/*GNP*	0.622	0.544	0.634	0.545	0.613	0.633
	(2.512)	(2.251)	(2.461)	(2.107)	(2.120)	(2.453)
Reserves/*GNP*	−1.155	−1.211	−1.079	−1.282	−0.995	−1.152
	(−2.164)	(−2.253)	(−1.632)	(−2.345)	(−1.412)	(−2.142)
Debt Service/Exports	0.426	0.567	0.440	0.441	0.386	0.353
	(1.688)	(2.344)	(1.797)	(1.749)	(1.400)	(1.458)
Loan Duration	−0.012	−0.011	−0.013	−0.014	−0.018	–
	(−0.648)	(−0.581)	(−1.719)	(−0.753)	(−0.953)	
Loan Value	−0.001	−0.001	−0.001	−0.001	−0.001	–
	(−1.340)	(−1.658)	(−1.269)	(−1.131)	(−1.500)	
Investment/*GNP*	−0.681	–	−0.756	−0.757	−0.624	−1.186
	(−1.991)		(−1.324)	(−1.318)	(−0.972)	(−2.266)
Current Account/*GNP*	0.435	–	0.387	0.487	0.365	–
	(1.966)	–	(0.970)	(2.131)	(0.863)	–
Growth	–	–	0.007	–	0.008	–
			(0.377)		(0.337)	
Imports/*GNP*	–	–	−0.004	–	0.007	–
			(−0.105)		(0.176)	
Government Expenditure/*GNP*	–	–	–	0.708	–	–
				(1.316)		
Income per Capita	–	–	–	–	−0.208	–
					(−0.532)	
Inflation	–	–	–	–	−0.008	–
					(−0.127)	
Reserves Variability	–	–	–	–	0.085	–
					(0.974)	
Rate of Devaluation	–	–	–	–	0.178	–
					(0.888)	
MSE	0.021	0.023	0.021	0.021	0.023	0.028

Note: Asymptotic *t*-statistics are shown in parentheses. *MSE* is the mean square error of the transformed regression.

ditional determinant of the desired level of international liquidity (see Eaton and Gersovitz, 1980). It is also interesting to note that the coefficient of the reserves ratio is quite high in absolute terms, exceeding in all cases the estimated value of the coefficient of the debt to *GNP* ratio.[9]

The coefficients of loan duration and loan value are negative, but insignificant. The coefficients of the imports-output ratio, growth, *GNP* per capita, variability of reserves, inflation, the government expenditure ratio, and the rate of devaluation are also insignificant.

In all regressions the estimated coefficient of the gross investment-*GDP* ratio was negative, as expected. Also in two of the five regressions where it was included, it was significant, suggesting that as has been indicated by Sachs (1981), Sachs and Cohen

[9]An interesting question is what will happen to the perceived probability of default if a country increases its foreign debt to finance the accumulation of international reserves. In this case the change in the (log of the) spread will be given by

$$d\log s=[\hat{\alpha}_1+\hat{\alpha}_2+\hat{\alpha}_3(\psi+i)/XR]\,d(DR),$$

where $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ are estimated regression coefficients of debt-*GNP*, reserves-*GNP*, and debt service-exports ratios, respectively; ψ is the fraction of the debt's principal that has to be amortized every year; i is the interest rate actually charged; XR is the exports-*GNP* ratio; and DR is the debt-*GNP* ratio. Computations reported in my earlier paper suggest that an increase in DR coupled with an equiproportional accumulation of reserves will tend to leave the spread unaffected.

(1982) and my earlier paper, a higher propensity to invest will tend to be associated with a lower perceived probability of default. The coefficient of the current account ratio is positive in the three regressions where it was included, being significant in two of the cases. This is a somewhat puzzling result, since it indicates that a lower deficit (or higher surplus) will result in a higher perceived probability of default and spread. The problem with this is that, with other things given—especially the investment ratio—a higher current account deficit means that the *same* investment is being financed with a higher proportion of foreign savings, and one would generally expect that in this case (i.e., lower domestic savings ratio) the perceived probability of default would be higher.

In all cases the estimated variance of the time-specific element $\hat{\sigma}_w^2$ exceeded the estimated country-specific variance $\hat{\sigma}_v^2$, indicating that during the period under consideration, differences across time in the country risk premium were more important than differences across country. This result is capturing the fact that throughout the period under consideration (1976–80) the level of world liquidity varied significantly (see my earlier paper). On the whole, however, the low value of the mean square error of the regressions (MSE) show a quite satisfactory fit.

In sum, the evidence presented in this section shows that during the recent past, lending behavior by international banks in Eurocurrency markets has taken into account (some of) the economic characteristics of borrowers. Even though some of the coefficients were sensitive to the specification of the estimated equations and were not always significant at the conventional levels, the general results regarding some of the most important variables are consistent with what was expected. Particularly important is the result of a significantly positive relation between the debt-GNP ratio and the spread. These results which contradict previous findings (i.e., Feder and Just, 1977b; Sachs, 1981) indicate that, as has been suggested by Harberger (1983), there are "externalities" in the process of LDC borrowing and that these could be dealt with by imposing an optimal tax on foreign borrowing in developing countries.

II. The Perceived Probabilities of Default

The econometric estimates reported in Table 1 can be used to compute the estimated banks' perceived probabilities of default as

$$(5) \quad p_{nt} = \frac{\exp\left\{\tilde{\alpha}_{0n} + \sum_{i=1}^{k} \hat{\alpha}_{in} y_{nti}\right\}}{1 + \exp\left\{\tilde{\alpha}_{0n} + \sum_{i=1}^{k} \hat{\alpha}_{in} y_{nti}\right\}},$$

where $\tilde{\alpha}_0 = \hat{\beta}_0 - k$ is the imputed value for α_0 in equation (4) (for $\hat{\beta}_0$ the estimated value of the constant in the regression analysis). Table 2 presents estimated probabilities of default for each country and each year, obtained from Table 1, equation (4.1).[10] A number of interesting characteristics of these probabilities can be observed. First, it can be seen that, within each year, there is a fairly wide variation in the perceived probability across countries. Second, for each country, these probabilities of default show some variation through time. For example, for the case of Venezuela the probability increases steadily between 1977 and 1980. While in 1976 Venezuela has the lowest estimated perceived probability of default, in 1980 this probability is around the middle of the distribution. For the case of Brazil, one of the countries that eventually ran into serious foreign debt problems, there is an increase in the perceived probability of default of approximately one full percentage point. Surprisingly, however, for Argentina, a country which in 1982 encountered serious financial difficulties, Table 2 estimates indicate that the perceived probability of default *declined* throughout the period. Also for the case of

[10]Each equation's estimates will generate different sets of perceived probabilities. However, the overall picture presented in Table 2 is not affected by the equation used to generate these probabilities. For further details on this and on the computation of these perceived probabilities of default, see my earlier paper.

TABLE 2—ESTIMATED PERCEIVED PROBABILITIES OF DEFAULT FROM TABLE 1, EQUATION (4.1) (Shown in Percent)

	1976	1977	1978	1979	1980
Greece	8.0	8.0	7.7	7.2	7.9
Portugal	8.4	8.3	8.5	8.9	8.6
Spain	7.8	8.1	8.1	7.8	7.9
Yugoslavia	7.6	7.6	7.8	7.0	7.4
Argentina	8.4	8.7	8.8	7.2	6.1
Brazil	8.9	8.8	8.9	9.1	9.6
Colombia	8.7	8.3	7.8	7.5	7.3
Ecuador	7.8	8.0	8.4	8.6	8.6
Mexico	9.9	10.1	10.5	10.5	9.2
Panama	10.4	11.6	11.9	11.3	11.5
Uruguay	10.6	10.2	10.5	8.6	8.5
Venezuela	5.9	5.8	6.3	7.1	7.9
Indonesia	9.6	9.6	9.3	8.4	8.7
Korea	8.9	8.5	7.7	7.7	8.5
Malaysia	7.0	7.4	7.0	7.1	6.3
Phillipines	7.4	7.9	7.7	7.7	7.7
Thailand	7.4	7.4	7.4	7.6	7.8
Ivory Coast	9.9	10.0	9.4	9.9	10.0
Morocco	8.0	8.1	9.4	9.2	10.3

Mexico, these results show a decline in the probability of default in 1980.

In sum, the computations presented in Tables 1 and 2 suggest that even as late as 1980 the international financial market had not predicted in any important way the future payment difficulties faced by Argentina, Mexico, and Uruguay.

III. Concluding Remarks

This paper has analyzed the relationship between foreign debt and default country risk in developing countries. The empirical analysis has used data on 727 public and publicly guaranteed loans granted to 19 *LDC*s during 1976 and 1980. The results obtained suggest that banks lending behavior has tended to consider (some of) the economic characteristics of countries when determining the spread they charge. However, the results also suggest that, at least during this period, banks might have overlooked some aspects of the developing countries' economies. One of the most interesting results obtained is the robust and significant positive relation (with a coefficient of approximately 0.6) between the log of the spread over *LIBOR* and the debt-output ratio. This relationship suggests that, as has been recently indicated by Harberger (1983), there are externalities in the process of *LDC* borrowing. These externalities could be dealt with by imposing an optimal external borrowing tax in these countries (Harberger, 1983).

REFERENCES

Anderson, Theodore W. and Hsiao, Cheng, "Estimation of Dynamic Models with Error Component," *Journal of the American Statistical Association*, September 1981, *76*, 598–606.

Beim, David, "Rescuing the *LDC*'s," *Foreign Affairs*, July 1977, *55*, 717–31.

Cline, William, *International Debt and the Stability of the World Economy*, Washington: Institute for International Economics, 1983.

Eaton, Jonathan, and Gersovitz, Mark, "*LDC* Participation in International Financial Markets: Debt and Reserves," *Journal of Development Economics*, March 1980, *7*, 3–21.

______ and ______, (1981a) "Debt With Potential Repudiation: Theoretical and Empirical Analysis," *Review of Economic Studies*, April 1981, *48*, 289–309.

______ and ______, (1981b) *Poor-Country*

Borrowing in Private Financial Markets and the Repudiation Issue, Princeton Studies in International Finance, No. 47, International Finance Section, Princeton University, June 1981.

Edwards, Sebastian, "*LDC*'s Foreign Borrowing and Default Risk: An Empirical Investigation, 1976–1980," Working Paper No. 1172, National Bureau of Economic Research, July 1983.

Feder, Gershon and Just, Richard E., (1977a) "A Study of Debt Servicing Capacity Applying Logit Analysis," *Journal of Development Economics*, March 1977, *4*, 25–38.

______ **and** ______, (1977b) "An Analysis of Credit Terms in the Eurodollar Market," *European Economic Review*, May 1977, *9*, 221–43.

______ **and** ______, "Optimal International Borrowing, Capital Allocation and Creditworthiness Control," *Kredit and Kapital*, No. 2, 1979, *12*, 207–20.

______ **and** ______, "A Model for Analyzing Lender's Perceived Risk," *Applied Economics*, June 1980, *12*, 125–44.

______ **and Ross, Knud,** "Risk Assessments and Risk Premiums in the Eurodollar Market," *Journal of Finance*, June 1982, *37*, 679–91.

Feldstein, Martin, "Coping with the International Debt Problem," unpublished manuscript, March 1983.

Folkerts-Landau, David F. I., "Credit Rationing and Debt Rescheduling in the International Bank Loan Market," unpublished paper, IMF, December 1982.

Frank, Charles R., Jr. and Cline, William R., "Measurement of Debt Servicing Capacity: An Application of Discriminant Analysis," *Journal of International Economics*, August 1971, *1*, 327–44.

Frenkel, Jacob, "International Liquidity and Monetary Control," Working Paper No. 1118, National Bureau of Economic Research, May 1983.

Fuller, Wayne and Batesse, George, "Estimation of Linear Models With Crossed Error Structure," *Journal of Econometrics*, February 1974, *2*, 67–78.

Harberger, Arnold C., "On Country Risk and the Social Cost of Foreign Borrowing by Developing Countries," unpublished manuscript, University of Chicago, 1976.

______, "Vignettes on the World Capital Market, *American Economic Review Proceedings*, May 1980, *70*, 331–37.

______, "Welfare Consequences of Capital Inflows," paper presented at World Bank Conference, Washington, October 1983.

______ **and Edwards, Sebastian,** "Lessons of Experience with Fixed Exchange Rates," in M. Gersovitz et al., eds., *The Theory and Experience of Economic Development*, London: Allen and Unwin, 1982, 183–93.

Kapur, Ishan, "An Analysis of the Supply of Eurocurrency Finance to Developing Countries," *Oxford Bulletin of Economics and Statistics*, August 1977, *37*, 171–88.

McDonald, Donogh C., "Debt Capacity and Developing Country Borrowing," *IMF Staff Papers*, December 1982, *29*, 603–46.

Sachs, Jeffrey D., "The Current Account and Macroeconomic Adjustment in the 1970s," *Brookings Papers on Economic Activity*, 1: 1981, 201–68.

______, "*LDC* Debt in the 1980s: Risk and Reforms," in Paul Wachtel, ed., *Crises in the Economic and Financial Structure*, Lexington: D. C. Heath, 1982, 197–243.

______, "Theoretical Issues in International Borrowing," National Bureau of Economics Research, Working Paper No. 1189, August 1983.

______ **and Cohen, Daniel,** "*LDC* Borrowing with Default Risk," Working Paper No. 925, National Bureau of Economic Research, July 1982.

Sargen, Nicholas, "Economic Indicators and Country Risk Appraisal," *Federal Reserve Bank of San Francisco Economic Review*, Fall 1977, 19–35.

Economist, March 5, 1983.

Euromoney, September 1978.

International Monetary Fund, *International Financial Statistics*, Washington, various issues.

Time, January 10, 1983.

Wall Street Journal, March 9, 1983.

World Bank, *Borrowing in International Capital Markets*, Washington, various issues.

______, *World Debt Tables*, Washington, various issues.

______, *World Development Report 1982*, New York: Oxford University Press, 1982.

______, *World Tables*, Baltimore: Johns Hopkins University Press, 1980.

Racial Discrimination in the Provision of Financial Services

By JAMES T. LINDLEY, EDWARD B. SELBY, JR., AND JOHN D. JACKSON*

The Equal Credit Opportunity Act of 1975 was amended in 1976 to expand the prohibition on discrimination in the extension of credit to include race, color, religion, national origin, and age. While studies have shown that differences exist between blacks and whites in capital accumulation (Henry Terrell, 1971) and in the use of financial services (Lindley-Selby, 1977), they have not concluded that the differences constituted racial discrimination in the supply of financial services. Evidence presented in support of the original Equal Credit Opportunity Act appears to have been statistically deficient in demonstrating discrimination based on sex. Richard Peterson concluded, "...that commercial banks did not systematically discriminate against potential borrowers based upon their sex before ECOA was passed" (1981, p. 560).

Testimony alleging racial discrimination in credit extension was given to Congress when it considered the 1976 amendment and to the Federal Reserve when it was in the process of promulgating Regulation B (Board of Governors, 1976, p. 243). Again, no statistical evidence supporting claims of racial discrimination was given. Despite the paucity of statistical evidence supporting the notion that financial institutions racially discriminate in the extension of credit, Congress acted as if such discrimination were pervasive.

The mood of Congress is reflected by the statement in the *Congressional Record* of Representative Frank Annunzio of Illinois:

> Mr. Speaker, there is far too much discrimination in the granting of credit in today's economic marketplace. And this discrimination must be eliminated.... [W]hy should a woman be denied the opportunity to open a charge account solely because she is black....
>
> In far too many cases, however, lenders merely reject or place great negative weight on factors such as age, race, color, creed, religion, or national origin of credit applicants.
>
> [1975, p. 3726]

The amendment was reported out of the Banking, Currency, and Housing Committee by a vote of 36 to 0 and was passed into law in 1976.

Review of the *Annual Report* of the Board of Governors of the Federal Reserve for the years 1976–81 reveals that between 2 and 4 percent of the complaints to the enforcing agencies alleged racial discrimination.[1] Such a small number of complaints casts further doubt on the pervasiveness of discrimination in the supply of financial services, particularly the supply of credit. And, in view of the significant administrative costs involved, the small number suggests a potential misallocation of resources.

Economic theory predicts that discrimination is a nonmaximizing activity in the face of competitive markets unless it is based on consumer prejudice. Consumer prejudice exists when, for example, white customers refuse to patronize an institution that accepts black customers (or conversely). If such a situation exists, the institution would have a

*Assistant Professor of Finance, University of Alabama, University, AL 35486; Professor of Banking and Finance, Department of Finance, University of Georgia, Athens, GA 30602; Associate Professor of Economics, Department of Economics, Louisiana Tech University, Ruston, LA 71272, respectively.

[1]The enforcing agencies and their responsibilities are as follows: 1) Federal Reserve, all member state banks; 2) Comptroller of the Currency, all national banks; 3) Federal Deposit Insurance Corporation, insured nonmember banks; 4) Federal Home Loan Bank Board, federally chartered savings and loans; 5) National Credit Union Administration, federal chartered credit unions; 6) Interstate Commerce Commission, regulated common carriers; 7) Civil Aeronautics Board, domestic and foreign air carriers; 8) Farm Credit Administration, federal land banks; 9) Securities and Exchange Commission, stock brokers and dealers; 10) Small Business Administration, small business investment companies; 11) Federal Trade Commission, all creditors not subject to the jurisdiction of the agencies mentioned above (Board of Governors, 1976, pp. 365–66).

profit-maximizing incentive to practice racial discrimination.

Discrimination based on the prejudice of the "manager" of the financial institution, however, can be maintained only as long as the firm is earning monopoly rents (Armen Alchian and Reuben Kessel, 1962).[2] If financial services are competitive, market forces will either push the discriminating institution out of business, or force it to accept black customers in order to stay in business. Thus, legislation prohibiting racial discrimination in the extension of credit is warranted if consumer prejudice exists, or if financial institutions are earning monopoly rents and are choosing to discriminate. Empirical attempts to measure the existence or absence of racial discrimination usually consist of observing either a significant coefficient on a race variable in a single estimated equation, or residual differences between black-white equations. Unfortunately, these procedures often shed little light on the discrimination issue.

In this paper an attempt is made to examine the discrimination issue by considering more than race variables or residuals. Since economic theory shows that the existence of discrimination is inconsistent with competitive markets except in special cases, the burden of proof lies with those who suggest the existence of discrimination. The task is to determine if there is evidence which supports a claim of discrimination. Rather than concentrating only on the race variable, the paper focuses upon the conditions which would be expected if discrimination in financial services existed, and then statistically tests to determine if those conditions exist.[3] In addition to the existence of a significant race variable, the conditions involve the degree of difference, if any, between discrimination in transaction services and in credit services, and the direction of change in the importance of the race variable over time.

The financial services offered by financial institutions can be divided into two types—transaction services (for example, checking accounts and savings accounts) and credit services (for example, loans and credit cards). Given that an institution has a profit-maximizing incentive to discriminate, it will tend to do so to a greater extent in the provision of credit than in transaction services. This result can be expected for two reasons. First, the role, length, and detail of the application for credit provide a much greater opportunity to discriminate. Second, since credit extension activities constitute the largest source of revenue to the institution, the institution's opportunity cost of not discriminating in the face of consumer prejudice would be considerably greater for credit extension than for transaction services.

In order to support the claim that discrimination exists in the supply of financial services, black-white differences in the use of financial services must be significant. Given that discrimination is more likely to exist where there is the greatest ability to discriminate and the largest return from practicing discrimination, these racial differences should be more pronounced in bank loans and credit cards (as opposed to savings and checking accounts). Finally, if discrimination

[2] Whether or not the American financial system is de facto competitive has been addressed in many studies. For the time period analyzed in this study, see John Scott (1977) and Federal Reserve Bank of Boston (1972). The question of the actual degree of competitiveness of the financial system is important to this study only in an *ex post* sense. A finding of discrimination gives implicit support to the hypothesis of a noncompetitive structure (and/or consumer prejudice), while a finding of no discrimination implicitly supports the converse hypothesis.

[3] Even in a fully and correctly specified model, inferring discrimination from the significance of the race variable alone can be misleading since it ignores possible interaction effects. Indeed, much of the discrimination literature employs a decomposition technique to account for this problem (Alan Blinder, 1973). The defect of Blinder's decomposition is that it assumes all interactions are supply determined. For a detailed discussion of the potential for incorrect inferences inherent in the rote application of this technique, see Jackson and Lindley (1983). As we argue later, significant interaction terms in this study are attributable to past discrimination in the supply of human capital and to consumer choice, and not to discrimination in the supply of financial services. Of course, if the model is not correctly specified, a significant race variable may be the result of a specification error (for example, omitted variables). Currently, the economics literature offers no tests for such specification errors that can be applied to probit estimation.

existed in the early 1970's, comparative data in later years should show a decrease in the importance of race as a determinant of financial usage, thereby reflecting legal and social changes which have occurred.

Given the above, there are three testable hypotheses. 1) Is the race varible significant? 2) Are racial differences more pronounced in credit variables than in transaction variables? 3) Is there a decline in the importance of race over time? If there is no statistical support for these three hypotheses, the case for discrimination in the extension of credit is extremely weak. Given that the burden of proof is to show discrimination, not the reverse, lack of statistical support for the three hypotheses presents the strongest case one can make from statistical inference—namely, there is not significant statistical support for the hypothesis that racial discrimination exists in the extension of credit. This, in turn, lends tacit support to the argument that racial discrimination in financial services is not profit maximizing and, therefore, is thwarted by the market.

I. Data Collection and Empirical Methodology

Under the assumption that an individual's intensity of preference for the use of a financial service is a function of his or her economic and demographic characteristics, we employ data collected from two temporally distinct surveys, one taken in 1971 and the other in 1979. Both surveys collected data from randomly selected households in the Atlanta, Georgia area using a cluster sampling technique. The 1971 sample contained 285 usable responses out of 321 interviews, and the 1979 sample contained 163 usable responses out of 181 interviews. The sample area contained, in addition to a number of competitive white-owned financial institutions, a black-owned bank and a black-owned savings and loan association, both long established.

The data contain both quantitative and qualitative variables. All of the qualitative variables are dichotomous. These variables are 1 or 0 for the use or nonuse of the particular financial service, 1 for car and house ownership and 0 for nonownership, and 1 for male-headed households and 0 for female-headed households. The remaining variables, household income, education level attained by the household head, age, and household population are of ratio strength.[4]

The intensity of an individual's preference for the use of a particular financial service is not directly measurable; all that can be observed is whether or not the individual uses the service. Thus, we are faced with the problem of estimating the probability of an individual employing a particular financial service while being interested specifically in the impact of race on this probability. We, therefore, employ the ordered N-Chotomous probit model developed by Richard McKelvey and William Zavoina (1975), and most recently applied in the finance literature by Richard Dietrich and Robert Kaplan (1982) and Kaplan and Gabriel Urwitz (1979). Since it was designed explicitly to handle the problems created by qualitative dependent variables, it is well suited to the analysis of the dichotomous data on financial services collected in the 1971 and 1979 surveys. Furthermore, since the parameter estimates are maximum likelihood estimates, they are known to be asymptotically normally distributed allowing for standard statistical tests to investigate the hypotheses posited earlier.

II. Impact of Race on Financial Variables

To determine the degree of statistical support for the hypotheses outlined above, the data were examined in three ways. First, the data were analyzed with conventional probit analysis incorporating a dummy variable for race as an explanatory variable. Next, the data were examined by estimating equations containing the race dummy and race interaction variables. Finally, the data were examined by estimating equations containing a time dummy and time interaction variables.

For the first two situations, six equations were estimated for the 1971 data and six equations for the 1979 data with the same set of independent variables. The dependent

[4] The data and detailed statistical results are available from the authors upon request.

TABLE 1—PROBIT COEFFICIENT ESTIMATES RELATING TO THE POTENTIAL EXISTENCE OF DISCRIMINATION

Financial Variable	Race Variable: No Interactions Included (1)	Race Variable: Race Interaction Included (2)	Race-Time Interaction Variable (3)
A. Least Likely to Discriminate			
Checking			
1979	−1.6234[c]	−1.9897[a]	
1971	−.6884[c]	−.5594	
Pooled			−1.0090[b]
Bank Saving			
1979	−.1980	.4189	
1971	.1447	−1.4481[c]	
Pooled			−.3604
Other Saving			
1979	−.4916[a]	−1.7842[c]	
1971	−.4663[b]	.0823	
Pooled			−.0046
B. Most Likely to Discriminate			
Bank Loan			
1979	−.2127	−.9085	
1971	−.4022[b]	.0418	
Pooled			.2958
Bank Card			
1979	−.2495	−.0764	
1971	−.5186[c]	−.9211	
Pooled			.2432
Other Card			
1979	.0377	.6662	
1971	−.2084	−.5504	
Pooled			.2029

Notes: Black = 1, white = 0. Although it is legitimate to draw inferences from these probit estimates concerning the sign and statistical significance of the effect of the variables on the probability of using a financial service, it is not legitimate to draw any inferences concerning their magnitude. (See Jackson and Richard Saba, 1981.)

[a]Significant at the 10 percent level.
[b]Significant at the 5 percent level.
[c]Significant at the 1 percent level.

variables in each of the six equations were checking accounts, bank savings accounts, other savings accounts, bank loans, bank credit cards, and other credit cards, respectively. The independent variables were income, head of household age, head of household sex, head of household education, population of the household, car ownership, house ownership, and race. All estimated equations converged and were significant at the 95 percent level using a *chi*-square test recommended by McKelvey and Zavoina. To create a time dummy variable and time interaction variables, the 1971 and 1979 data were pooled resulting in a set of six equations with the above dependent and independent variables.

The dependent variables were divided into two categories: "Most Likely to Discriminate," which includes credit-based services (bank loan, bank credit card, and other credit card); and "Least Likely to Discriminate," which includes transaction services (checking, bank saving, and other saving). The rationale for this distinction has been discussed earlier.

Table 1 contains a summary of the results. Column 1 displays the results of conventional probit analysis and shows that race is a significant variable for some financial services indicating that there are black-white differences in the use of these services. However, only two of the six race variables in group B are significant, while four of the six

variables in group A are significant. This is exactly opposite the expected results if the hypothesis is true that discrimination is more likely in credit extension. Some support for the proposition that alleged discrimination has decreased over time exists in that the two significant race variables in the category Most Likely to Discriminate were significant in 1971 and became insignificant in 1979 while there was no change in the Least Likely to Discriminate category. It is tempting to conclude that the 1976 legislation was both needed and somewhat effective in reducing racial discrimination in the extension of credit services.

However, a potential problem with the above analysis, which is not readily apparent, is that rather than a restriction in the supply of a financial service causing the race variable to be significant or insignificant, the cause may have been the impact of socioeconomic variables on the probability of using a financial service. To determine the extent of this effect, race interaction variables were created by multiplying the race dummy by each of the included socioeconomic variables and the equations reestimated.[5] The coefficients on these interaction variables reflect black-white differences in the response of the probability of using a particular financial service to *ceteris paribus* changes in the explanatory variables. For example, a significant race-income interaction coefficient in the analysis of checking accounts indicates that the change in the probability of having a checking account caused by a change in income differs significantly for blacks vs. whites.

Statistically significant interaction terms are the result of cultural differences in the way changes in socioeconomic variables affect preferences for financial services. They may very well be the result of past discrimination in income earning potential or education availability, but they cannot be interpreted as the result of contemporary discrimination in the supply of financial services. This latter source of discrimination is indicated again by the race dummy. The inclusion of race interaction variables allows a clearer indication of the source of black-white differentials in the use of financial services and produces results which are more appropriate to answering the question of restriction in the supply of financial services based on race.[6,7]

The results, displayed in Table 1, column 2, show that while there are significant race variables, they are all in the financial services in the Least Likely to Discriminate category. Overall, the support for the theory of the existence of discrimination in the supply of financial services is relatively weak because race is not significant for the financial services in which financial institutions would be most likely to discriminate.

To further test support for the existence of discrimination in the supply of financial services, the impact of time on the significance of the race variable was explored. Pooling the data, a time dummy ($=0$ for

[5]The use of dummy variables and dummy interaction terms to detect intercept and slope differences, respectively, between groups of data is discussed in J. Johnston (1972, pp. 176–86). To our knowledge, this procedure has not been applied in probit analysis. Nonetheless, the extension is straightforward. Consider the change in the probability of a particular event (R_m) caused by a change in the lth explanatory variable for the ith observation. If the ith individual is white (race = 0), then $\partial(R_{m,1})/\partial X_l = \hat{\beta}_l\ [\cdot]$ where $\hat{\beta}_l$ is the estimated probit coefficient of X_l and $[\cdot]$ is $[f(\Sigma\hat{\beta}_j X_{ji} - \hat{\mu}_{j-1}) - f(\Sigma\hat{\beta}_j X_{ji} - \hat{\mu}_m)]$, f being the standard normal density function. If the individual is black (race = 1), then $\partial(R_{m,i})/\partial X_l = (\hat{\beta}_l + \hat{\beta}_p)\ [\cdot]$ where $\hat{\beta}_p$ is the estimated probit coefficient of the variable created as the product of the race dummy and X_l. The difference in these probability changes is $\hat{\beta}_p[\cdot]$. At least for the dichotomous case, this difference will be significant if β_p is significant. Thus, a statistically significant interaction coefficient indicates a statistically significant difference in the way the probability of a particular event is affected by changes in a given explanatory variable for blacks vs. whites.

[6]If significant interaction variables cause the race dummy to become insignificant, the implication is that all observed black-white differences stem from differences in the way changes in socioeconomic variables affect preferences for the use of financial services. If significant interaction variables cause an insignificant race dummy to become significant, the possibility that offsetting effects from changes in the socioeconomic variables obscured the detection of actual racial differences in the previous analysis is suggested.

[7]The equations were first run with all of the variables including all interaction variables. However, not all of the equations converged due partly to the large number of variables relative to the number of observations. The equations were rerun with only the significant variables with their corresponding interaction variables and the variable race. All of the equations then converged.

1971; =1 for 1979) and corresponding interaction terms were created.[8] Of importance is the sign and significance of the coefficient on the race-time interaction variable. A negative significant coefficient indicates an increase in the importance of race as a variable over time, while a positive significant interaction coefficient shows a decrease.[9] These estimates are summarized in column 3.

The results are striking in that while the signs for the services in the category Most Likely to Discriminate support the claim that there has been a decrease in the importance of race, none of the variables are significant. This result suggests that whether or not discrimination was a problem in 1971, the passage of the 1976 legislation produced no significant change in the situation (since the black-white difference in the probability of employing any of these services did not change significantly from 1971 to 1979). Coupled with our prior results incorporating race interaction terms, this further suggests that the legislation was not only ineffective, but also unnecessary. The variables in the category Least Likely to Discriminate have the opposite sign indicating race became more significant over time. The financial service, checking accounts, is the only one for which the race-time variable was found to be significant, but with a puzzling sign which indicates racial differences in the probability of having a checking account *increased* over time.

Because checking is a service in which banks are least likely to discriminate, it is rather doubtful that this result is attributable to an increase in discrimination against blacks in the supply of checking accounts. Even if this is the case, however, our conclusions concerning the efficacy of the 1976 legislation are unaltered. A more likely explanation for this result is differential growth in the demand for cash balances for transaction purposes. The demand by blacks grew less rapidly than that of whites.

III. Conclusions

An amendment was adopted in 1976 adding race to the list of characteristics which are prohibited as a basis for discrimination in the extension of credit. While pervasive racial discrimination in the extension of credit was alleged, little statistical evidence has been provided to support the allegation. In this paper, the discrimination issue has been examined by determining whether or not conditions existed which would be expected if the alleged discrimination existed.

There is no significant statistical support for the hypothesis that discrimination exists in the extension of credit. Thus, we cannot reject the argument that racial discrimination in financial services is not profit maximizing and, therefore, is thwarted by the market. The results in this paper are consistent with those of George Benston, Dan Horsky, and Martin Weingartner (1978) with respect to redlining. They did not find evidence to support a driving need for antidiscrimination regulation in the extension of mortgages. The results are also consistent with those of Peterson who found no evidence that lenders discriminated on the basis of sex prior to the passage of the Equal Credit Opportunity Act.

[8] In the eight-year period separating the two samples, changes took place which could affect the usage of financial services. For example, the relaxation of Regulation Q could have caused transaction services to become more costly. So long as the effects of these changes are race neutral, the effects can be read from the significance of the time dummy alone. If the changes impacted one race more than the other, the effect could show up also in the race-time interaction coefficient. This is one possible explanation (although in our view not a likely one) for the negative significant race-time interaction term for checking accounts in the third column of Table 1.

[9] This result may not be readily apparent to the reader. It is perhaps most easily understood by realizing that the race-time interaction coefficient indicates the difference in the black-white difference in the probability of a person using a given financial service in 1971 vs. 1979. If this difference is to have decreased over time (i.e., if discrimination has declined over time), the coefficient must be positive.

REFERENCES

Alchian, Armen A. and Kessel, Reuben A., "Competition, Monopoly, and the Pursuit of Pecuniary Gain," in *Aspects of Labor Economics*, NBER Universities-Conference Series, No. 13, New York: Arno Press,

1962, 157–75.

Benston, George J., Horsky, Dan and Weingartner, H. Martin, *An Empirical Study of Mortgage Redlining*, Monograph Series in Finance and Economics, No. 1978–5, New York University, 1978.

Blinder, Alan S., "Wage Discrimination: Reduced Form and Structural Estimates," *Journal of Human Resources*, Fall 1973, *8* 436–55.

Dietrich, J. Richard and Kaplan, Robert S., "Empirical Analysis of the Commercial Loan Classification Decision," *Accounting Review*, January 1982, *62*, 18–38.

Jackson, John D. and Lindley, James T., "On the Empirical Measurement of Discrimination: A Statistical Test and A Caveat," Working Paper No. 42, University of Alabama, 1983.

______ **and Saba, Richard,** "On the Testing of Behavioral Hypothesis in N-Chotomous Probit Analysis," paper presented at the Southwestern Federation of Administrative Disciplines, New Orleans, 1981.

Johnston, J., *Econometric Methods*, 2nd ed., New York: McGraw-Hill, 1972.

Kaplan, Robert S. and Urwitz, Gabriel, "Statistical Models of Bond Ratings: A Methodological Inquiry," *Journal of Business*, April 1979, *52*, 231–61.

Lindley, James T. and Selby, Edward B., Jr., "Racial Differences in the Use of Financial Services," *American Journal of Economics and Sociology*, October 1977, *36*, 393–99.

McKelvey, Richard D. and Zavoina, William, "A Statistical Model for the Analysis of Ordinal Level Dependent Variables," *Journal of Mathematical Sociology*, No. 1, 1975, *4*, 103–20.

Peterson, Richard L., "An Investigation of Sex Discrimination in Commercial Banks' Direct Consumer Lending," *Bell Journal of Economics*, Autumn 1981, *12*, 547–61.

Scott, John Troy, "Price and Nonprice Competition in Banking Markets," Research Report 62, Federal Reserve Bank of Boston, 1977.

Terrell, Henry S., "Wealth Accumulation of Black and White Families: The Empirical Evidence," *Journal of Finance*, May 1971, *26*, 363–77.

Board of Governors of the Federal Reserve, *Annual Report*, 1976–81.

Federal Reserve Bank of Boston, *Policies for a More Competitive Financial System*, Conference Series No. 8, June, 1972.

U.S. Congressional Record, 94th Cong., 1st Session, Vol. 121, Part 3, Washington: USGPO, 1975.

Lucas on the Quantity Theory: Hypothesis Testing without Theory

By CHARLES H. WHITEMAN*

In a recent paper, Robert Lucas presents "Two Illustrations of the Quantity Theory of Money." His intriguing illustrations are of the two central quantity-theoretic propositions: a change in the rate of growth in the money supply results in equal changes in the inflation rate and the nominal rate of interest. Lucas finds these propositions compelling because they "...possess a combination of theoretical coherence and empirical verification shared by no other propositions in monetary economics" (1980, p. 1005). For empirical verification he points, for instance, to Robert Vogel's (1974) study of Latin American inflation, wherein average money growth-inflation rate pairs for sixteen countries approximately lie on a 45° line. By "theoretical coherence," Lucas means that the quantity theory propositions arise as necessary conditions for equilibrium in well-posed models of economies in which agents optimize and markets clear. The propositions hold, for instance, in the model economies studied by Miguel Sidrauski (1967a, b).

Standing in contrast to the quantity theory propositions is the Mundell-Tobin effect of an increase in the rate of growth of the money stock: the resulting inflation lowers the yield on money holdings and induces an asset shift to real capital. Thus, in the presence of this effect, an increase in the growth rate of the money supply causes a less than equiproportional increase in the nominal rate of interest. Though there is, perhaps, a dearth of empirical evidence on this effect, it too possesses a good deal of theoretical coherence. For instance, in the overlapping generations model of fiat money studied by Neil Wallace (1980), John Bryant and Wallace (1979), and others, the Mundell-Tobin effect is the primary one.

Apparently, the resolution of the conflict between these propositions requires no more than a standard exercise in empirical macroeconomics. Indeed, Lucas writes

> In the absence of the kind of decisive natural experiment used by Vogel, one could in principle test the neoclassical laws by deriving their implications for the parameters of a structural econometric model. This course, while attractive in theory...is in practice a difficult one, since it involves nesting the two hypotheses in question within a complex maintained hypothesis, which must be accepted as valid in order to carry out the test. [1980, p. 1007]

Eschewing such structural estimation, Lucas proceeds to investigate the quantity theory propositions using a technique intended to uncover long-run empirical regularities but which is not guided by economic theory. He defends this strategy, arguing

> The virtue of relatively atheoretical tests, such as carried out by Vogel, is that they correspond to our theoretically based intuition that the quantity theoretic laws are consistent with a wide variety of possible structures. If so, it would be desirable to test them independently and then, if confirmed, to *impose* them in constructing particular structural models, rather than to proceed in the reverse direction. [p. 1007]

If one accepts Lucas's technique, the results apparently provide some evidence against the Mundell-Tobin effect. However, without any theoretical guidance, many readers may find the results difficult to interpret.

Some light may be shed on these results by

*Univerisity of Iowa, Iowa City, IA 52242. I am indebted to Thomas Sargent and Neil Wallace and members of the Macroeconomic Workshop at the University of Virginia for helpful discussions on the issues discussed in this paper.

taking a variant of Lucas's first suggestion seriously; that is, by using Lucas's techniques to analyze the long-run properties of a widely used, theoretically defensible, econometrically tractable structural model. The model, or acceptable "complex maintained hypothesis," to be used for this purpose is Lucas's (1975) equilibrium model of the business cycle. The model is attractive because it isolates the Mundell-Tobin effect in a single parameter.

Section I establishes that Lucas's technique is a robust method for calculating the sums of lag coefficients from long distributed lag regressions of inflation and nominal interest rates on the growth rate of the money supply. The model is presented and interpreted in Section II, and the relevant sums of distributed lag coefficients are calculated in Section III. The results of this exercise indicate that the Lucas results are not necessarily evidence against the Mundell-Tobin effect. In fact, his results can be interpreted as evidence in favor of the effect. That this interpretation is at variance with Lucas's is a simple implication of the Lucas (1976) critique of econometric policy evaluation: Lucas's technique does uncover some information about the presence (or absence) of the Mundell-Tobin effect, but it is contaminated by information about the constraints faced by the agents of the economy. The paper is thus a Lucas-critique-based admonition: beware of hypothesis testing without theory.

I. The Data Processing Techniques

Lucas's procedure involved two steps. First, he applied a two-sided filter to time-series for the rate of money creation (μ_t), the rate of inflation (π_t), and the nominal rate of interest (ρ_t). For instance, he replaced the time-series $\{\mu_t\}$ by

$$(1)\quad \{\mu_t(\beta)\} = \left\{\left(\frac{1-\beta}{1+\beta}\right)\sum_{k=-\infty}^{\infty} \beta^{|k|}\mu_{t-k}\right\},$$

with $0<\beta<1$.[1] Second, he generated plots of the pairs $(\pi_t(\beta), \mu_t(\beta))$ and $(\rho_t(\beta), \mu_t(\beta))$ for various values of β, and found that for β near 1.0, the filtered observations approximately fell on a 45° line; that is, his graphical method discovered R^2, a_1, and a_2 near 1.0 in the regressions

$$(2a)\quad \pi_t(\beta) = a_1\mu_t(\beta)+\varepsilon_{1t} \quad E\varepsilon_{1t}\mu_t(\beta)=0,$$

$$(2b)\quad \rho_t(\beta) = a_2\mu_t(\beta)+\varepsilon_{2t} \quad E\varepsilon_{2t}\mu_t(\beta)=0.$$

In order to judge his method and results, some results from projection theory will prove useful.

Consider the projection of ρ_t on the entire μ_t process:

$$(3)\quad \rho_t = \sum_{k=-\infty}^{\infty} \gamma_k\mu_{t-k} + \eta_t,$$

where η_t satisfies $E\mu_{t-k}\eta_t=0$ for all k. Defining $S_{\rho\rho}(e^{-iw})$ and $S_{\mu\mu}(e^{-iw})$ as the spectral densities of $\{\rho_t\}$ and $\{\mu_t\}$, the coefficients γ_k can be recovered by (inverse) Fourier transforming

$$\gamma(e^{-iw}) = S_{\rho\mu}(e^{-iw})/S_{\mu\mu}(e^{-iw}),$$

where $S_{\rho\mu}(e^{-iw})$ is the cross spectrum of (ρ_t, μ_t).

In practice, (3) is impossible to estimate: it contains an infinite number of parameters. Clearly, some restriction of $\{\gamma_k\}$ is necessary. There are many such restrictions; for example, truncation: $\gamma_k=0$ for all $k>N_1>0$ and $\gamma_k=0$ for all $k<N_2\le 0$. Let $\{\gamma_k'\}$ be a restricted version of $\{\gamma_k\}$ such that $\{\gamma_k'\}\ne\{\gamma_k\}$ for any such restriction. The restricted regression to be implemented is

$$(4)\quad \rho_t = \sum_{k=-\infty}^{\infty} \gamma_k'\mu_{t-k} + u_t.$$

Least squares chooses $\{\hat{\gamma}_k'\}$ to minimize the sum of squared residuals $\sum_{t=0}^{T} u_t^2$. But this sum of squared residuals is proportional to the area under the spectrum of u_t, $S_{uu}(e^{-iw})$. Substituting (3) into (4) and rearranging yields

$$(5)\quad u_t = \sum_{k=-\infty}^{\infty} (\gamma_k-\gamma_k')\mu_{t-k} + \eta_t.$$

[1]The filter was truncated in a particular way at the end of the sample. See Lucas (1980, fn. 5, p. 1007).

From (5), the spectrum of u_t is

$$S_{uu}(e^{-iw}) = |\gamma(e^{-iw}) - \gamma'(e^{-iw})|^2 S_{\mu\mu}(e^{-iw}) + \sigma_\eta^2/2\Pi$$

where $\gamma(e^{-iw}) = \Sigma_{k=-\infty}^{\infty}\gamma_k e^{-iwk}$. Hence, least squares chooses $\{\gamma_k'\}$ to minimize

$$(6)\quad \int_{-\Pi}^{\Pi} |\gamma(e^{-iw}) - \gamma'(e^{-iw})|^2 S_{\mu\mu}(e^{-iw})\, dw$$

which is Christopher Sims's (1972a) approximation error formula. Clearly, $\{\gamma_k'\}$ best approximates $\{\gamma_k\}$ at frequencies w where $S_{\mu\mu}(e^{-iw})$ is the largest. Since (angular frequency w) $\times$ (length—in time—of sine wave cycle) $= 2\Pi$, good "long-run" approximations require relatively large power in $\{u_t\}$ at low frequencies. If the long-run properties of $\{\gamma_k\}$ are important, it may be useful to filter μ_t with a filter which has large power at low frequencies w. Such a filter is said to be a "low pass" filter.

Lucas notes that the filter in (1) has Fourier transform

$$(7)\quad \beta(e^{-iw}) = (1-\beta)^2/(1+\beta^2 - 2\beta\cos w)$$

which, for large β, concentrates most of its power at low frequencies. His graphical method forces the a_2 of (2b) to approximate $\{\gamma_k\}_{-\infty}^{\infty}$ of (3), so surely approximation error is present, and Sims's formula (6) is useful: Lucas's methods amount to choosing a_2 to minimize

$$\int_{-\Pi}^{\Pi} |\gamma(e^{-iw}) - a_2|^2 S_{\mu\mu(\beta)}(e^{-iw})\, dw.$$

Thus, the function $\gamma(e^{-iw})$ is being approximated by a function which is constant for $w \in (-\Pi, \Pi)$. When the spectrum of the unfiltered money creation process, $S_{\mu\mu(0)}(e^{-iw})$, has power at low frequencies, the spectrum of the filtered money creation process, $S_{\mu\mu(\beta)}(e^{-iw})$, will, for β near 1.0, have most of its power at low frequencies. This means that large weight is placed on getting the approximation "correct" for very small w. Thus, when β is large, a_2 approximates $\gamma(e^{-i0}) = \gamma(1)$, the sum of the lag coefficients in the regression of ρ_t on $\{\dots, \mu_{t+1}, \mu_t, \mu_{t-1}, \dots\}$. Moreover, this will be true regardless of the time-series properties of $\{\rho_t\}$ and $\{\mu_t\}$.

Of course, these considerations apply equally to Lucas's graphs of $\pi_t(\beta)$ against $\mu_t(\beta)$. His graphical exhibition of exponentially smoothed time series is simply a method of estimating sums of distributed lag coefficients, albeit an extremely roundabout one.[2]

[2]Why not estimate the sum by $\gamma'(1)$, as Terence Mills (1982) does? The reason, as Sims (1972a) demonstrated, is that even when $\{\gamma_k'\}$ provides a "good fit," $\Sigma_{k=-\infty}^{\infty}\gamma_k'$ may be a very poor estimate of $\Sigma_{k=-\infty}^{\infty}\gamma_k$. This is apparent from (6): the sum of lag coefficients is *one* point ($\gamma(1)$) on the *function* $\gamma(e^{-iw})$, and because the "interval" $w=0$ has no measure, $\gamma'(1)$ can be changed arbitrarily without affecting the value of (6)—the "fit." To take an example, consider the $\gamma'(e^{-iw})$ function estimated by Mills (equation (24), p. 1167):

$$\gamma'(e^{-iw}) = \underset{(0.003)}{0.02}\left(1 - \underset{(0.2)}{2.52}\, e^{-iw} + \underset{(0.4)}{2.10}\, e^{-2iw} - \underset{(0.21)}{0.57}\, e^{-3iw}\right)^{-1}$$

where the numbers in parentheses are the coefficient standard errors. Though Mills reports $\gamma'(1) = 1.4$, because he gives coefficients with only 3-digit accuracy, it appears that $\gamma'(1) = 2.0$. Notice, though, that, holding the others constant, as the last coefficient varies from -0.56 to -0.58 (a change which apparently amounts to less than 10 percent of one standard error), $\gamma'(1)$ increases from unity to plus infinity. Thus changes in estimated coefficients which do not materially affect the fit are associated with huge changes in the estimated sum of lag coefficients. This is not quite fair because Mills (probably) used an unconditional maximum-likelihood technique which rules out unit roots, and my example is strictly valid only if the coefficients were estimated using conditional maximum likelihood (conditional on data from the first three quarters of 1957) which does not. In particular, the estimated coefficients will be correlated and a change in one will induce changes in the others in such a way that the sum in the denominator will not be identically zero—though it can easily be made close enough to establish my point. At minimum, it is clear that the reported standard errors are misleading. Technically, this occurs because the sum of coefficients is not in general continuous with respect to the "goodness of fit" distance measure. Lucas's method docs not share this problem because it employs a sum of coefficients function (the constant a_2) that is

However the sum is estimated, there remains the question of what it means: Lucas takes a_1 and a_2 near unity as evidence in favor of the quantity theory and against the Mundell-Tobin effect. Would this evidence be sufficient to convince a nonbeliever? The answer to this question involves the use of a fully articulated artificial economy which can be used to perform population versions of Lucas's calculations. This artificial economy, Lucas's (1975) equilibrium model of the business cycle, is both a convenient and proper framework for studying the Mundell-Tobin effect.

II. The Model

Let y_t denote the log of output at time t; k_t, the log of the capital stock at t; r_t, the log of the real gross return to capital at t; m_t the log of the money stock at t; and p_t, the log of the price level at t. Output is given by

$$(8) \qquad y_t = \delta_0' + \delta_1' k_t \qquad \delta_1' > 0,$$

while the real return to capital follows

$$(9) \qquad r_t = \delta_0 - \delta_1 k_t \qquad \delta_1 > 0.$$

Expressions (8) and (9) can be thought of as arising from a setting in which markets are competitive and workers supply their services inelastically to firms whose technology is Cobb-Douglas in capital and labor.

There are two assets in the model, real capital and real balances. The demand for capital is governed by

$$(10) \quad k_{t+1} = \alpha_0' + \alpha_1 E_t r_{t+1} + \alpha_2 (E_t p_{t+1} - p_t) + \alpha_3 k_t,$$

where $\alpha_1 > \alpha_2 \geq 0$, $\alpha_3 \in (0,1)$, and $E_t x$ is the linear least squares projection of x on information known at date t. The information set includes current (date t) and past values of all the variables listed above. The demand for real balances is given by

$$(11) \quad m_t - p_t = \beta_0' - \beta_1 E_t r_{t+1} - \beta_2 (E_t p_{t+1} - p_t) + \beta_3 k_t$$

with $\beta_2 > \beta_1 > 0$ and $\beta_3 \in (0,1)$.

The capital stock is fixed at date t and has a one-period gestation; that is, date t output from the production process (8) is divided among new capital (date $t+1$), consumption, and government expenditures which are financed totally by money creation. The sequence of money supplies $\{m_t\}_{t=-\infty}^{\infty}$ is taken to be a stochastic process with moving average representation

$$(12) \qquad m_t = \sum_{k=0}^{\infty} A_k e_{t-k},$$

where $e_t = m_t - E(m_t | m_{t-1}, m_{t-2}, \ldots)$: $\{e_t\}$ is fundamental for $\{m_t\}$. Equation (12) expresses m_t as the convolution of the two sequences $\{A_k\}$ and $\{\varepsilon_t\}$. With the understanding that $A(L) = \sum_{k=0}^{\infty} A_k L^k$, where the lag operator L is defined by $L^n x_t \equiv x_{t-n}$, (12) can also be written $m_t = A(L) e_t$.

The interesting system dynamics are completely determined by (10), (11), and (12). The demand for capital, given in (10), is positively related to the expected return to capital, $E_t r_{t+1}$, and inversely related to the expected return to money holdings $p_t - E_t p_{t+1}$. Similarly, the demand for end-of-period real balances, given in (11), is positively related to the own-expected return and negatively related to the expected return to real capital.

The magnitude of the Mundell-Tobin effect is measured by the parameter α_2. When $\alpha_2 = 0$, the effect is absent. In this case, because $E_t r_{t+1}$ is a deterministic function of k_{t+1}, the demand for capital is independent of the expected inflation rate. This demand, using (9), is described by

$$(13) \quad k_{t+1} = \frac{\alpha_0' + \alpha_1 \delta_0}{1 + \alpha_1 \delta_1 - \alpha_3} + c_k \left[\frac{\alpha_3}{1 + \alpha_1 \delta_1} \right]^t,$$

by construction continuous with respect to the least squares metric. In so doing, it sacrifices accuracy in estimating individual lag coefficients in favor of accuracy concerning the sum. In addition, the "controlled" approximation error which is so useful for estimating the sum of coefficients destroys any hope of obtaining the standard error of that estimate.

which converges to $\bar{k}=(\alpha_0'+\alpha_1\delta_0)/(1+\alpha_1\delta_1-\alpha_3)$ for any finite arbitrary constant c_k. Clearly, monetary disturbances do not affect the steady-state capital stock.

The demand for real balances when $\alpha_2=0$ is given by

$$(14)\quad m_t-p_t=\beta_0'-\beta_1\delta_0-\beta_2(E_tp_{t+1}-p_t) + (\beta_1\delta_1+\beta_3L)k_{t+1},$$

where k_{t+1} is given by (13). This is a version of Phillip Cagan's (1956) portfolio balance schedule, albeit one driven by the deterministic sequence k_t. After means have been removed, the stochastic steady state for the price level is given by the solution of (14) for p_t:[3]

$$p_t=\frac{1}{1+\beta_2}\sum_{k=0}^{\infty}\left[\frac{\beta_2}{1+\beta_2}\right]^k E_tm_{t+k}.$$

The sum of the lead weights in this expression is 1.0. Thus, in a sense, an increase in the *log* of the money stock m_t leads, eventually, to an equal increase in the *log* of the price level p_t. A similar result holds for the nominal rate of interest, since the real rate, given by (13) and (9), is fixed in steady state. Thus, when $\alpha_2=0$, the Mundell-Tobin effect is absent, and the model exhibits versions of the two quantity theory propositions.

When $\alpha_2>0$, the model exhibits the Mundell-Tobin effect. Monetary shocks will affect the price level, the capital stock, and, by (9), the real rate of interest. The quantity theory results are overturned when $\alpha_2\neq 0$, but a precise description of how this happens must be postponed until a solution (reduced-form) of the model is derived.

In solving the model, it will be convenient to view equations (10), (11), and (12) as restrictions on the moving average representation of the variables of the model. Since k_{t+1} is determined at time t, the three variables are m_t, k_{t+1}, and p_t. Also, it will be assumed that the pair (k_{t+1},p_t) fails to Granger-cause m_t. This assumption, which rules out feedback from economic conditions to the money supply, makes it possible to concentrate on the quantity theory and the Mundell-Tobin effect. By Sims's Theorem 1 (1972b, p. 544), the Granger-ordering assumption means that the moving average representation for the indeterministic parts of m_t, k_{t+1}, and p_t can be written

$$(15)\quad \begin{pmatrix} m_t\\ k_{t+1}\\ p_t\end{pmatrix} = \begin{pmatrix} A(L) & 0 & 0\\ B(L) & D(L) & G(L)\\ C(L) & F(L) & H(L)\end{pmatrix}\begin{pmatrix} e_t\\ w_t\\ v_t\end{pmatrix},$$

where A, B, C, D, etc. are polynomials in the lag operator, for example, $C(L)=C_0+C_1L+C_2L^2+\cdots$, and e_t, w_t, and v_t are jointly fundamental for m_t, k_{t+1}, and p_t.

The restrictions (9)–(12) must hold for all realizations of e_t, w_t, and v_t. But since the coefficients in A, B, C, etc. depend only on the distributions of e_t, w_t, and v_t, (9)–(12) impose restrictions on the coefficients of these polynomials. In addition, since D, F, G, and H do not enter into the projections to be calculated in the next section, the only important restrictions are those on A, B, and C.

Given an $A(L)$—a "money supply rule," the assumption of rational expectations uniquely determines $B(L)$ and $C(L)$ as[4]

$$(16a)\quad B(L)=\frac{\alpha_2}{\Delta(1-\lambda_1L)(1-\lambda_2L)} \times\{(1-\lambda_2^{-1})A(\lambda_2^{-1})-(1-L)A(L)\},$$

$$(16b)\quad C(L)=\frac{1}{\Delta(1-\lambda_1L)(1-\lambda_2L)} \times\{\Delta(1-\lambda_2^{-1})A(\lambda_2^{-1})(1-aL) -(1+\alpha_1\delta_1-\alpha_3L)LA(L)\},$$

[3] This solution is obtained, for example, by Thomas Sargent (1979, p. 269).

[4] The details of these calculations can be found in my 1981 dissertation, ch. 7. In particular, Δ, λ_1, and λ_2 are defined implicitly by $\Delta(1-\lambda_1L)(1-\lambda_2L)=(1+\alpha_1\delta_1-\alpha_3L)(\beta_2-(1+\beta_2)L)-\alpha_2(1-L)(\beta_1\delta_1+\beta_3L)$. A general description of the methods employed is given in my 1983 study.

where $\Delta > 0$, $0 < \lambda_1 < 1 < \lambda_2$, and $a = (\alpha_2\beta_3 + \alpha_3\beta_2)/\Delta$. The projections of the interest rate and the rate of inflation on the rate of money growth are easily computed using (16). The sums of the distributed lag coefficients from these projections are calculated in the next section.

III. Sums of Distributed Lag Coefficients

In Lucas's model, the nominal interest rate, which is the real rate plus the expected inflation rate, is given by $r_t + (E_t p_{t+1} - p_t)$; the inflation rate by $(1-L)p_t$; and the rate of growth of the money supply by $(1-L)m_t$. The reproduction of Lucas's calculations thus requires computation of the sums of the distributed lag coefficients in the regressions

$$\text{(17a)}\quad (1-L)p_t = \sum_{k=-\infty}^{\infty} \gamma_k^{\pi}(1-L)m_{t-k} + \eta_t^{\pi}$$

$$\text{(17b)}\quad r_t + (E_t p_{t+1} - p_t) = \sum_{k=-\infty}^{\infty} \gamma_k^{\rho}(1-L)m_{t-k} + \eta_t^{\rho}$$

where $E(1-L)m_t\eta_{t-s}^{\rho} = E(1-L)m_t\eta_{t-s}^{\pi} = 0$ for all s.

The Z-transform of the distributed lag coefficients $\{\gamma_k^{\pi}\}_{k=-\infty}^{\infty}$ in (17a) can be calculated using (15) and methods described by Thomas Sargent:[5]

$$\gamma^{\pi}(Z) = \frac{(1-Z)C(Z)(1-Z^{-1})A(Z^{-1})}{(1-Z)A(Z)(1-Z^{-1})A(Z^{-1})} = C(Z)/A(Z).$$

Then (16) may be used to calculate the sum of lag coefficients $\gamma^{\pi}(1)$:

$$\text{(18)}\quad \gamma^{\pi}(1) = \frac{1}{\Delta(1-\lambda_1)(1-\lambda_2)} \times \left\{\Delta(1-a)(1-\lambda_2^{-1})A(\lambda_2^{-1})A(1)^{-1} - (1+\alpha_1\delta_1 - \alpha_3)\right\}.$$

[5] Ordinarily, the variance of e_t would appear in these expressions. But the calculations of the previous section require that (20) be normalized by taking $\mathrm{var}(e_t u_t v_t)' = I$.

Suppose that there is no Mundell-Tobin effect; that is, that $\alpha_2 = 0$. Then, the sum of the lag coefficients reduces to

$$\gamma^{\pi}(1) = 1 - \frac{\beta_2}{1+\beta_2}A\left(\frac{\beta_2}{1+\beta_2}\right)A(1)^{-1}.$$

Since $\beta_2 > 0$, a necessary condition for Lucas's result ($\gamma^{\pi}(1) = 1$) is $A(1)^{-1} = 0$—a condition which will hold if and only if the money supply has a random walk component, the condition employed by Lucas (1972, 1975). But when this condition holds and $\alpha_2 \neq 0$, $\gamma^{\pi}(1) = -(1+\alpha_1\delta_1 - \alpha_3)/\Delta(1-\lambda_1)(1-\lambda_2)$ which, as it turns out, is precisely equal to 1. Thus, as expected, Lucas's first result is essentially independent of the magnitude of the Mundell-Tobin effect.

The calculation of the sum of lag coefficients in (17b) is somewhat more complicated. First, write $A^*(L) = (1-L)A(L)$. Second, notice that, using (15), $E_t p_{t+1} - p_t = p_{t+1} - p_t - C_0 e_{t+1}$ plus terms in $\{v_t\}$ and $\{w_t\}$. Third, by the zero-mean version of (9), $r_t = -\delta_1 k_{t+1} = -\delta_1 B(L)e_t$ plus terms in $\{v_t\}$ and $\{w_t\}$. Thus the coefficients $\{\gamma_k^{\rho}\}_{k=-\infty}^{\infty}$ in (17b) are those in the least squares projection of $(-\delta_1 B(L) + L^{-1}(1-L)C(L) - C_0 L^{-1})e_t$ on $\{(1-L)m_t\}_{t=-\infty}^{\infty}$. Using the linearity of the projection operator, and (again) using techniques described by Sargent,

$$\gamma^{\rho}(Z) = -\delta_1 B(Z)/(1-Z)A(Z) + Z^{-1}\gamma^{\pi}(Z) - C_0 Z^{-1}/(1-Z)A(Z).$$

Using (16), $\gamma^{\rho}(Z)$ can be written

$$\gamma^{\rho}(Z) = \frac{-\delta_1\alpha_2}{\Delta(1-\lambda_1 Z)(1-\lambda_2 Z)} \times \frac{A^*(\lambda_2^{-1}) - A^*(Z)}{A^*(Z)} + \frac{\gamma^{\pi}(Z)}{Z} - \frac{A^*(\lambda_2^{-1})}{A^*(Z)}$$

since $C_0 = (1-\lambda_2^{-1})A(\lambda_2^{-1}) = A^*(\lambda_2^{-1})$. The

sum of lag coefficients $\Sigma_{k=-\infty}^{\infty}\gamma_k^\rho$ is then

$$\gamma^\rho(1)=\frac{-\delta_1\alpha_2}{\Delta(1-\lambda_1)(1-\lambda_2)}\times\left\{\frac{A^*(\lambda_2^{-1})-A^*(1)}{A^*(1)}\right\}+\gamma^\pi(1)-\frac{A^*(\lambda_2^{-1})}{A^*(1)}.$$

At this stage, two questions naturally arise. From the point of view taken by Lucas, the more interesting one is: given $\gamma^\rho(1)=1$, what can one infer about α_2? A less interesting but still relevant question is: does $\alpha_2=0$ imply $\gamma^\rho(1)=1$? It is convenient but innocuous to condition the answers to these questions on Lucas's first result that $\gamma^\pi(1)=1$. Then $\gamma^\rho(1)=1$ implies

$$\frac{\alpha_2\delta_1}{\Delta(1-\lambda_1)(1-\lambda_2)}\times\{1-A^*(\lambda_2^{-1})/A^*(1)\}=A^*(\lambda_2^{-1})/A^*(1), \tag{19}$$

a restriction involving all of the parameters of the model as well as those governing the evolution of the money supply. This restriction in turn implies $\alpha_2=0$ only if $A^*(\lambda_2^{-1})/A^*(1)=0$—which occurs only if either the growth rate of the money supply has a random walk component ($A^*(1)^{-1}=0$), or it fails to have an autoregressive representation ($A^*(\lambda_2^{-1})=0$). Given Terence Mills's results (p. 1165), neither of the two conditions seems plausible. Moreover, Lucas's own results provide evidence against $A^*(1)^{-1}=0$: under this condition, the spectrum of the money creation process already has infinite power at zero frequency, and the use of the filter in (1) with successively larger β's (which enhances low frequency power) should have no effect on the estimated a_1 and a_2. A glance at Lucas's figures confirms that the coefficients do indeed vary with β.[6]

To answer the converse question, let $\alpha_2=0$. Then $\gamma^\rho(1)=1-A^*(\lambda_2^{-1})/A^*(1)$, so the sum of lag coefficients is unity only under the same conditions on the money supply derived above.

These calculations indicate that within the context of Lucas's (1975) model, his (1980) empirical results depend crucially on the structure of the money supply process. If either of two (unlikely) conditions on the money supply process holds, the sum of lag coefficients in a regression of the nominal interest rate on the rate of money creation will equal unity if and only if there is no Mundell-Tobin effect. But the most plausible interpretation of the results taken as a whole —a_1 and a_2 which vary with β and approach unity as β does—is that (19) holds with $\alpha_2\neq 0$; that there *is* a Mundell-Tobin effect. That this interpretation differs from Lucas's is a consequence of the Lucas (1976) critique: reduced-form correlations (which he used in his calculations) are complicated functions of the structure of the economy as well as of the laws of motion of processes individuals care about. As such, they are unlikely to provide much evidence about the "true" model of the economy unless one already has it very much in mind.[7]

IV. Conclusion

Lucas's (1980) calculations summarize interesting long-run relationships in a novel way. But calculations like these are bound to depend on more than just behavioral parameters. The analysis above used this simple implication of the Lucas (1976) critique and a version of his (1975) equilibrium model of the business cycle to show that his (1980) calculations do not properly pit the quantity theory propositions against the Mundell-Tobin effect. Indeed, it seems appropriate to withhold judgment about the relative amounts of "empirical verification" the two classes of propositions possess. Certainly, enough doubt has been cast to refrain from adopting Lucas's suggestion to *impose* the

[6]For the skeptical, I replicated Lucas's (π,μ) calculations with $\beta=0$, 0.5, 0.9, and 0.95. The associated a_1s were 0.02, 0.08, 0.87, and 0.99.

[7]Bennett McCallum (1984) illustrates this point in a different setting.

quantity theory propositions in constructing particular structural models.

REFERENCES

Bryant, John and Wallace, Neil, "The Inefficiency of Interest-Bearing National Debt," *Journal of Political Economy*, April 1979, *87*, 365–81.

Cagan, Phillip, "The Monetary Dynamics of Hyperinflation," in Milton Friedman, ed., *Studies in the Quantity Theory of Money*, Chicago: University of Chicago Press, 1956.

Fischer, Stanley, "Anticipations and the Nonneutrality of Money," *Journal of Political Economy*, April 1979, *87*, 225–52.

Granger, Clive W. J., "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, July 1969, *37*, 424–38.

Hansen, Lars Peter, and Sargent, Thomas J., "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, February 1980, *2*, 7–46.

Lucas, Robert E., Jr., "Expectations and the Neutrality of Money," *Journal of Economic Theory*, April 1972, *4*, 103–24.

______, "An Equilibrium Model of the Business Cycle," *Journal of Political Economy*, December 1975, *83*, 1113–44.

______, "Econometric Policy Evaluation: A Critique," in Karl Brunner and Alan Meltzer, eds., *The Phillips Curve and Labor Markets*, Vol. 1, Carnegie-Rochester Conference Series on Public Policy, *Journal of Monetary Economics*, Suppl. 1976, 19–46.

______, "Two Illustrations of the Quantity Theory of Money," *American Economic Review*, December 1980, *70*, 1005–14.

McCallum, Bennett T., "On Low-Frequency Estimates of 'Long-Run' Relationships in Macroeconomics," *Journal of Monetary Economics*, forthcoming 1984.

Mills, Terence C., "Signal Extraction and Two Illustrations of the Quantity Theory," *American Economic Review*, December 1982, *72*, 1162–68.

Mundell, Robert A., "Inflation and Real Interest," *Journal of Political Economy*, June 1963, *71*, 280–83.

Muth, John F., "Optimal Properties of Exponentially Weighted Forecasts," *Journal of the American Statistical Association*, June 1960, *55*, 299–306.

Sargent, Thomas J., *Macroeconomic Theory*, New York: Academic Press, 1979.

Sidrauski, Miguel, (1967a) "Rational Choice and Patterns of Growth in a Monetary Economy," *American Economic Review Proceedings*, May 1967, *57*, 534–44.

______, (1967b) "Inflation and Economic Growth," *Econometrica*, December 1967, *33*, 671–84.

Sims, Christopher A., (1972a) "The Role of Approximate Prior Restrictions in Distributed Lag Estimation," *Journal of the American Statistical Association*, March 1972, *67*, 169–75.

______, (1972b) "Money, Income, and Causality," *American Economic Review*, September 1972, *62*, 540–52.

Tobin, James, "Money and Economic Growth," *Econometrica*, October 1965, *33*, 671–84.

Vogel, Robert C., "The Dynamics of Inflation in Latin America, 1950–1969," *American Economic Review*, March 1974, *64*, 102–14.

Wallace, Neil, "The Overlapping Generations Model of Fiat Money," in John Kareken and Neil Wallace, eds., *Models of Monetary Economies*, Minneapolis: Federal Reserve Bank of Minneapolis, 1980.

Whiteman, Charles H., "Moving Average Representations in Rational Expectations Models," unpublished doctoral dissertation, University of Minnesota, 1981.

______, *Linear Rational Expectations Models: A User's Guide*, Minneapolis: University of Minnesota Press, 1983.

The Firm in Short-Run Industry Equilibrium: Comment

By MICHAEL BRAULKE*

It has long been recognized that the traditional law of supply and demand valid for the isolated competitive firm need not necessarily hold when considering the individual firm within the greater context of its industry. Indeed, a change in a certain input or output price will in general affect all input and output decisions, and these reactions can be expected to trigger other price changes at the industry level in those markets where the industry faces less than infinitely elastic supply or demand. The initial response of the individual firm to the original price change may thus be altered by its response to these consequent price changes and perhaps even be completely reversed.

In his 1982 paper, Ronald Heiner has shown for the special yet important case of an infinitely elastic supply in all factor markets that at least the law of demand will nevertheless hold in the short run for the entire industry, provided that demand for its output is normal in the usual sense of a downward-sloping schedule. This reassuring result is noteworthy in itself, but it also commands admiration because it represents one of the few cases where an ambiguity at the micro level is resolved at the macro level by aggregation. Heiner goes even further in the characterization of the industry's short-run response by saying that it will be weaker than the response that would arise if demand in the output market were infinitely elastic and conversely stronger than the response that would result if demand in the output market were entirely inelastic (pp. 555–56). Yet in his formal analysis Heiner proves a somewhat different (and weaker) version of this characterization by showing that the industry's short-run response must be stronger than the response that would be observed if the output of every firm were fixed.

It is the purpose of this comment to show that Heiner's intuitively more appealing original contention indeed holds and in fact characterizes the industry's short-run response more precisely than the version that was actually proven. At the same time, and perhaps more interestingly, I want to generalize Heiner's findings by showing that they in essence still apply if the industry also faces in some of its input markets a less than infinitely elastic (but normal) supply. This generalization is obviously relevant for all those industries that are the sole users, or at least major users of certain inputs and hence not necessarily "small" in relation to the size of these input markets. By generalizing Heiner's notation, I use a framework that allows us to deal with an arbitrary number of inputs and outputs and, in particular, with an arbitrary number and combination of markets in which the industry is confronted with a less than infinitely elastic demand or supply.

I. The Industry's Short-Run Response

In order to make sure which side of the market we are talking about when speaking of demand and supply at the industry level, let us first agree on some terminology. In what follows, *aggregate demand and supply* denotes the aggregate demand for the industry's outputs and the aggregate supply of its inputs whereas the *industry's supply and demand* means its aggregate supply of outputs and its aggregate demand for inputs.

Consider an industry characterized in the short run by a fixed number, F, of competitive but not necessarily identical firms. Denoting the vector of output and input quantities of firm j by $z^j = (x^j, y^j)$, where x^j and y^j are subvectors of arbitrary length, I will use the common convention that outputs are positive, inputs negative. Let subvector x comprise all outputs and inputs for which the corresponding aggregate demand and supply is infinitely elastic, so that subvector y comprises all the remaining outputs and inputs that are characterized by less than infinitely elastic aggregate demand and

*Department of Economics, Universität Osnabrück, F.R.G. I thank Ronald Heiner for helpful suggestions.

supply. Further, let the price vector $p = (\alpha, \beta)$ be ordered and partitioned conformably with the quantity vector z. Then only the prices α are exogenous to the industry whereas the prices β are determined endogenously at the industry level by a market-clearing mechanism.

The main object of interest is the individual firm's actual short-run supply and demand of outputs and inputs, x, as a function of the prices α which incorporates the repercussions from simultaneous adjustments in the prices β which always have to clear their respective input or output markets. Denote this actual short-run supply and demand by $\tilde{x}^j$ and define for comparative purposes the additional hypothetical concepts: the firm's short-run supply and demand $\hat{x}^j$ that would be observed if aggregate demand and supply in the y markets were in fact infinitely elastic; its short-run supply and demand $\check{x}^j$ that would result if this aggregate demand and supply were conversely entirely inelastic; and finally the firm's short-run supply and demand $\bar{x}^j$ for the hypothetical case that its supply and demand y were held fixed at a just-binding level.

It remains to be clarified as to what is meant by normal conditions in the less than infinitely elastic y markets. Let $R(\beta)$ denote the aggregate demands and supplies corresponding to the quantity vector y. They are assumed to have the usual slopes with respect to their own-prices (demands and supplies negatively and positively sloped, respectively). If these markets are not interconnected through the β prices, the matrix of derivatives R_β is diagonal, with all the (nonpositive) own-price effects on the main diagonal. If these markets are interconnected, it will be assumed that the own-price effects outweigh the cross effects as implied in the usual semidefiniteness conditions for the optimizing firm. Thus, in general, it will be assumed $R(\beta)$ is "normal" in the sense of R_β being negative semidefinite, denoted by $R_\beta \leq 0$.[1,2]

Given these concepts and definitions, the following theorem (the proof follows in Section II) that is essentially due to Heiner can now be stated:[3]

THEOREM 1: *Assume that aggregate demand and supply, $R(\beta)$, in the less than infinitely elastic markets is normal in the sense $R_\beta \leq 0$. The industry's actual short-run supply and demand, $\tilde{X}(\alpha)$, will then obey the traditional law of supply and demand. More precisely, at the industry level the relations*

$$\hat{X}_\alpha \geq \tilde{X}_\alpha \geq \check{X}_\alpha \geq \bar{X}_\alpha \geq 0$$

will hold. However, a corresponding sequence of relations need not necessarily hold for individual firms within the industry.

The theorem says that industry level effects in the y markets (whose aggregate demands and supplies $R(\beta)$ may not be infinitely elastic) will not upset the traditional law of supply and demand in the remaining x markets so long as $R(\beta)$ is normal. That is, even when we allow industry level repercussions in some input or output markets to alter their market-clearing prices, the industry will still obey the law of supply and demand in reacting to price changes in the remaining markets. It will weaken though the industry's actual short-run response as compared to the situation where aggregate demand and supply in all markets is infinitely elastic ($\hat{X}_\alpha \geq \tilde{X}_\alpha$). On the other hand, the industry's actual short-run response will in general still be stronger than the response in the extreme situation where aggregate demand and supply in the y markets is

[1] The symbols "≥ 0" and "≤ 0" are used as a shorthand expression for "positive" or "negative semidefinite." Accordingly, $C \geq D$ means $C - D \geq 0$.

[2] To further illustrate the $R_\beta \leq 0$ assumption, consider an equilibrium situation described by $R(\beta^*) = Y$, where Y is the industry's supply and demand. Suppose there is an exogenous increase (decline) in the industry's supply (or demand) so that $dY \geq 0$. The consequent change in equilibrium prices is then characterized by $R_\beta d\beta^* = dY$. Multiplying from the left by $d\beta^{*\prime}$ we have $d\beta^{*\prime} R_\beta d\beta^* = d\beta^{*\prime} dY$. The condition $R_\beta \leq 0$ guarantees that the weighted sum of price changes, $d\beta^{*\prime} dY$, is nonpositive. Thus, for example, an exogenous increase in industry supply could never lead to an overall increase in equilibrium prices.

[3] Capital letters are used to denote summation over all F firms in the industry. Thus $X = \Sigma x, Y = \Sigma y$, etc.

entirely inelastic ($\tilde{X}_\alpha \geq \check{X}_\alpha$). Finally, this latter bound characterizes the industry's actual short-run response more precisely than the response $\overline{X}_\alpha$ which would occur if all firms in the industry had to produce and use fixed amounts of the y quantities.

II. Proof of Theorem 1

Since I follow essentially Heiner's basic line of reasoning, I can be brief in the preparatory steps of the proof.

Consider first the regular profit-maximizing choice of the individual firm, denoted by $\hat{z}^j(p)$, and remember that, as Eugene Silberberg (1974) has shown very elegantly, the associated response matrix

$$(1) \qquad \hat{z}_p^j = \begin{bmatrix} \hat{x}_\alpha^j & \hat{x}_\beta^j \\ \hat{y}_\alpha^j & \hat{y}_\beta^j \end{bmatrix}$$

is positive semidefinite and symmetric. This implies in particular $\hat{x}_\alpha^j \geq 0$, $\hat{y}_\beta^j \geq 0$ and the useful symmetry relation

$$(2) \qquad \hat{x}_\beta^j = \hat{y}_\alpha^{j\prime}.$$

It implies furthermore that the corresponding industry response matrix

$$(3) \qquad \hat{Z}_p = \Sigma_j \hat{z}_p^j = \begin{bmatrix} \hat{X}_\alpha & \hat{X}_\beta \\ \hat{Y}_\alpha & \hat{Y}_\beta \end{bmatrix}$$

is also positive semidefinite and symmetric.

Assume now for a moment that the inputs and outputs belonging to group y are fixed for each firm at a certain level y^j. Denoting the corresponding profit-maximizing choice by $\bar{z}^j(\alpha, y^j)$, the indentity $\hat{z}^j(p) = \bar{z}^j(\alpha, \hat{y}^j(p))$ must clearly hold which implies in particular

$$(4) \qquad \hat{x}^j(\alpha,\beta) = \bar{x}^j(\alpha, \hat{y}^j(\alpha,\beta)).$$

This in turn yields the identities[4]

$$(5a) \qquad \hat{x}_\beta = \bar{x}_y \hat{y}_\beta,$$

$$(5b) \qquad \hat{x}_\alpha = \bar{x}_\alpha + \bar{x}_y \hat{y}_\alpha = \bar{x}_\alpha + \bar{x}_y \hat{y}_\beta \bar{x}_y^\prime,$$

where (2) and (5a) were used in the derivation of the last equality. Since $\hat{y}_\beta \geq 0$, (5b) is nothing but an alternative statement of the well-known LeChâtelier phenomenon which says that a firm's unrestricted price response will not be weaker than its response under (just binding) quantity rationing in some markets, that is, $\hat{x}_\alpha \geq \bar{x}_\alpha$. Note that $\bar{x}_\alpha \geq 0$ by standard results.

Consider next the market-clearing mechanism that equilibrates the industry's supply and demand with aggregate demand and supply so that

$$(6) \qquad \hat{Y}(\alpha,\beta) = R(\beta)$$

holds. As Heiner pointed out, equation (6) can be viewed as an implicit representation of the functional relationship, $B(\alpha)$, between the equilibrium prices β and the exogenous prices α. Substituting $B(\alpha)$ into (6) and differentiating with respect to α leads then to the key relation

$$(7) \qquad \hat{Y}_\alpha + \hat{Y}_\beta B_\alpha = R_\beta B_\alpha.$$

Consider instead the hypothetical case where aggregate demand and supply in these particular markets is entirely inelastic, that is, where it is not given by $R(\beta)$ but equal to a constant vector $\overline{R}$. Equilibrium then normally yields a different solution, say $\check{B}(\alpha)$, for the corresponding market-clearing prices β. In contradistinction to (7), this solution must obviously satisfy

$$(8) \qquad \hat{Y}_\alpha + \hat{Y}_\beta \check{B}_\alpha = 0.$$

Substituting $B(\alpha)$ for β in $\hat{x}(\alpha,\beta)$ gives then

$$(9) \qquad \tilde{x}(\alpha) = \hat{x}(\alpha, B(\alpha))$$

whereas substituting $\check{B}(\alpha)$ for β gives

$$(10) \qquad \check{x}(\alpha) = \hat{x}(\alpha, \check{B}(\alpha)).$$

Since the other two concepts, $\hat{x}(\alpha)$ and $\bar{x}(\alpha)$, are already defined, the proof of the theorem is as follows.

[4]To simplify notation, I neglect from now on the firm-specific superscript.

PROOF:

(i) $\hat{X}_\alpha \geq \tilde{X}_\alpha$. Since by (9) and (2) $\tilde{x}_\alpha = \hat{x}_\alpha + \hat{y}_\alpha' B_\alpha$ and accordingly $\tilde{X}_\alpha = \hat{X}_\alpha + \hat{Y}_\alpha' B_\alpha$ it suffices to show that $\hat{Y}_\alpha' B_\alpha$ is negative semidefinite. Multiplying (7) from left with B_α', transposing and rearranging terms gives $\hat{Y}_\alpha' B_\alpha = B_\alpha'(R_\beta' - \hat{Y}_\beta) B_\alpha$, which is clearly negative semidefinite given $R_\beta \leq 0$.

(ii) $\tilde{X}_\alpha \geq \check{X}_\alpha$. By (9), (10), and (2), we have $\tilde{X}_\alpha = \hat{X}_\alpha + \hat{Y}_\alpha' B_\alpha$ and $\check{X}_\alpha = \hat{X}_\alpha + \hat{Y}_\alpha' \check{B}_\alpha$. Hence, we must show that $\hat{Y}_\alpha'(B_\alpha - \check{B}_\alpha)$ is positive semidefinite. Note first that by (7) and (8), $R_\beta B_\alpha = \hat{Y}_\beta(B_\alpha - \check{B}_\alpha)$ and that by (8), $\hat{Y}_\alpha' = -\check{B}_\alpha' \hat{Y}_\beta$. Using these results, consider the quadratic form $(B_\alpha' - \check{B}_\alpha')\hat{Y}_\beta(B_\alpha - \check{B}_\alpha) = B_\alpha' \hat{Y}_\beta(B_\alpha - \check{B}_\alpha) - \check{B}_\alpha' \hat{Y}_\beta(B_\alpha - \check{B}_\alpha) = B_\alpha' R_\beta B_\alpha + \hat{Y}_\alpha'(B_\alpha - \check{B}_\alpha) \geq 0$. Thus, $\hat{Y}_\alpha'(B_\alpha - \check{B}_\alpha) \geq 0$ must hold since $R_\beta \leq 0$.

(iii) $\check{X}_\alpha \geq \bar{X}_\alpha$. By (10), (5a) and (5b) we may write $\check{x}_\alpha = \bar{x}_\alpha + (\bar{x}_y \hat{y}_\beta \bar{x}_y' + \bar{x}_y \hat{y}_\beta \check{B}_\alpha)$. It suffices therefore to show that the expression in brackets, when summed over all firms in the industry, is positive semidefinite. Consider the quadratic form $(\bar{x}_y + \check{B}_\alpha')\hat{y}_\beta(\bar{x}_y' + \check{B}_\alpha) = (\bar{x}_y \hat{y}_\beta \bar{x}_y' + \bar{x}_y \hat{y}_\beta \check{B}_\alpha) + \check{B}_\alpha'(\hat{y}_\beta \bar{x}_y' + \hat{y}_\beta \check{B}_\alpha) \geq 0$. Note that by (5a) and (2) the last expression in parentheses can be written in the form $(\hat{y}_\alpha + \hat{y}_\beta \check{B}_\alpha)$. In view of (8), this last expression vanishes when summed over all firms. Hence, $\Sigma(\bar{x}_y \hat{y}_\beta \bar{x}_y' + \bar{x}_y \hat{y}_\beta \check{B}_\alpha)$ must indeed be positive semidefinite.

(iv) $\bar{X}_\alpha \geq 0$. Note that $\bar{X}_\alpha = \Sigma \bar{x}_\alpha$.

REFERENCES

Heiner, Ronald A., "Theory of the Firm in Short-Run Industry Equilibrium," *American Economic Review*, June 1982, *72*, 555–62.

Silberberg, Eugene, "A Revision of Comparative Statics in Economics, or, How to Do Comparative Statics on the Back of an Envelope," *Journal of Economic Theory*, February 1974, *7*, 159–72.

The Firm in Short-Run Industry Equilibrium: Further Comment

By RONALD A. HEINER*

Michael Braulke presents an elegant generalization of my analysis on industry factor demand. In support of this work, let me suggest a way of allowing the aggregate demands and supplies in the y markets to be fully interrelated through both the α and β prices, rather than just the β prices alone.[1]

The objective is to show that a subset of arbitrarily selected inputs or outputs, x, will obey the law of demand or supply for the entire industry. Now allow the aggregate demands-supplies of the other y markets to be interrelated through all prices (α, β), denoted $R(\alpha, \beta)$. Consider a particular y^i market. As firms react to a change in α, both the aggregate supply-demand R^i and industry demand-supply Y^i will shift in response to equilibrium changes in the other β prices besides β^i.

We can trace out an adjusted aggregate supply-demand which shows how R^i reacts to β^i with these equilibrium changes in the other β prices already built in; that is, to maintain equilibrium in the other markets as R^i reacts to β^i. Let $\tilde{R}^i(\beta^i)$ denote this adjusted aggregate supply-demand for y^i. Similar adjustment will generate the whole vector $\tilde{R}(\beta)$, where each subfunction, $\tilde{R}^i(\beta^i)$, depends only on its own-price. This implies that $\tilde{R}_\beta$ is a diagonal matrix, where the elements on the main diagonal are the slopes of the adjusted supplies and demands for changes to their own-prices. Then assume $\tilde{R}_\beta \leq 0$, which now means the demands-supplies $\tilde{R}$ have the usual slopes when adjusted to preserve equilibrium in the other y markets as α changes.[2]

Use the implicit price equation $\beta = B(\alpha)$ to form the modified market-clearing equation, $\tilde{Y}(\alpha) \equiv \tilde{R}[B(\alpha)]$, where $\tilde{Y}(\alpha) = \hat{Y}[\alpha, B(\alpha)]$. This immediately implies

$$(1) \qquad B'_\alpha \tilde{Y}_\alpha = B'_\alpha \tilde{R}_\beta B_\alpha \leq 0.$$

Eugene Silberberg's (1974) method is used to set up the following indirect Lagrangian for each firm,

$$(2) \qquad \phi(x, y, \alpha) = \alpha \tilde{x}(\alpha) + B'(\alpha)\tilde{y}(\alpha) - [\alpha x + B'(\alpha) y].$$

Use the envelope theorem and simplify to obtain, $\phi_{\alpha\alpha} = \tilde{x}_\alpha + B'_\alpha \tilde{y}_\alpha \geq 0$. Summing over firms and then substituting in (1) implies, $\tilde{X}_\alpha + B'_\alpha \tilde{R}_\beta B_\alpha \geq 0$. This in turn implies from (1), $\tilde{X}_\alpha \geq 0$; and thus we are done.

*Department of Economics, Brigham Young University, Provo, UT 84602.

[1]The same terminology and notation as Braulke's paper are used.

[2]Thus, the direct effect of each β^i on R^i obeys the law of supply or demand, and these direct effects are sufficient to counteract repercussions from maintaining equilibrium in other markets.

REFERENCES

Braulke, Michael, "The Firm in Short-Run Industry Equilibrium: Comment," *American Economic Review*, September 1984, *74*, 750–53.

Silberberg, Eugene, "A Revision of Comparative Statics in Economics, or, How to Do Comparative Statics on the Back of an Envelope," *Journal of Economic Theory*, February 1974, *7*, 159–72.

Union Wage Policies: Comment

By PAUL G. CHAPMAN AND MALCOLM R. FISHER*

A recent paper in this *Review* (1981) by Ian McDonald and Robert Solow (M-S) provided an account of wage stickiness throughout different phases of the business cycle. Our comment is concerned with the M-S conclusion and suggests why wage flexibility may still persist. First, it is argued that demand pressure may influence the "fairness criteria" which are used to select any (efficient) wage-employment solution. Second, in cases where inefficient solutions occur, it is argued that the assumption that demand changes are isoelastic cannot be maintained as a reason for sticky wages if there is no basis for this type of demand change. Third, in cases where demand changes are isoelastic, the critical M-S assumption is that union leaders give equal weighting to the preferences of both the employed and unemployed union members. It is shown below how union preferences for employed union members increase the likelihood of wage flexibility, at least for upturns in the business cycle. Fourth, the aggregation problems attached to the use of the M-S model for economywide predictions suggest further uncertainties in their results.

The wage bargaining model developed by McDonald and Solow is based on three relationships. First, the labor demand curve defines the employment levels which maximize profits given a particular wage (the isoprofit functions representing various levels of profit will be at their maximum on L_D, the labor demand curve). Second, the contract curve represents the locus of efficient points where the union indifference curves have the same slope as the isoprofit curves. Third, a solution on the contract curve can be derived subject to an assumption about the relative dominance of either the employer or the union. If the union is dominant, a fair division of net revenue will be obtained so that wages form $100k$ percent of revenue. If the employer is dominant, the solution will lie SW from this point on the contract curve. These relationships are represented in Figure 1. In both cases, the solution lies on the intersection of the contract curve (upward sloping) and the equity curve (downward sloping).

McDonald and Solow show that a change in demand has an indeterminate effect on real wages (p. 903). This follows from the counteracting influences of demand on the equity locus and the contract curve. The demand increases push the curves to the right ensuring a rise in employment but allowing wages to move in either direction or remain stable. It follows that any factors which reinforce the shift in the equity locus increase the likelihood of a positive relation between wages and demand changes. Therefore if k, the fair share of wages in revenue, is a positive function of demand, which would seem plausible, then the case for wage flexibility is strengthened.[1]

A further problem arises if union dominance is reduced to the extent that L_D becomes the effective constraint; the union can choose the wage rate but the employer chooses the employment rate. In this case, the bargaining solution lies on L_D and not on the efficiency frontier. This type of solution may conflict with Paretian efficiency criteria, but corresponds more closely with labor market realities where unions may not

*Department of Economics, University of Dundee, DD1 4HN Scotland, and Australian Graduate School of Management, University of New South Wales, P.O. Box 1, Kensington, NSW, respectively.

[1]This can be readily shown in a formal way for the M-S model by writing $k = k(B)$ and deriving the sign of dW/dB as $-[(1 - LR_L/R)R_{LB} + LR_{LL}R_B/R + R_{LL}k_B]$ where B represents demand and R represents net revenue; R_L, R_{LL}, R_B and k_B represent the relevant partial derivatives.

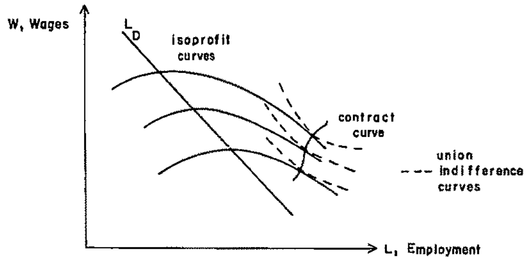


FIGURE 1

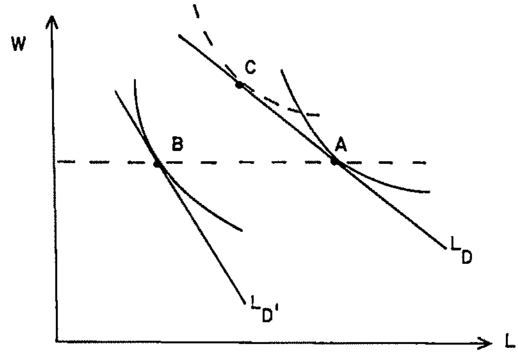


FIGURE 2

wish to be seen to bargain over employment in terms of higher wages. In these circumstances, the power locus is clearly irrelevant to the solution; M-S conclude in this case that wage stickiness occurs when $L = G(W, B)$ can be written as $G(W, B) = BG(W)$ (p. 899). This follows from equating the slopes of the union indifference curve and the L_D curve. The special case which results in sticky wages is where the demand schedules are isoelastic for different levels of B, that is, the elasticity remains constant at a given wage for the various L_D schedules as B changes. It follows that the union locates a wage on the L_D curve with a given elasticity and as B changes will stick to this rate.

However, there is no reason to expect isoelastic shifts in the demand curve for labor; therefore this special case cannot be a general explanation for wage stickiness. It is, of course, a formal statement of the conditions necessary for wage stickiness (for inefficient solutions). It can also be noted that $G = BG(W)$ implies that L_D tilts around a point on the wage axis as B changes. In cases where parallel shifts in L_D occur (dG/dB is constant), the elasticity of demand falls at a given wage and the preferred solution is a higher wage and higher employment as L_D shifts to the right. Clearly the elasticity question is an empirical issue and there is no theoretical case in favor of or against sticky wages. Furthermore we might expect the overall demand for labor to vary throughout the business cycle even where individual firms (or industries) have isoelastic demand schedules. The *weighted* average aggregate demand curve will have a variable elasticity if the individual firms are affected differently by the variations in demand.

A third deficiency in the sticky wages hypothesis can arise for the inefficient case even though demand shifts are isoelastic. An important assumption of the model is that there are a fixed number of union members and, therefore, in the general case there will be some unemployed union members. This raises the possibility that the union may adopt an asymmetric policy towards demand changes. If demand falls the relevant group of workers is the currently employed labor force and the union maximizes the utility function where, as M-S show, wages will be rigid for the isoelastic case. This is shown as A to B in Figure 2. If the reverse change in demand is assumed ($L_{D'}$ to L_D), the union may not move from B to A because two groups of members are involved (currently employed and unemployed). The two groups have different interests and if the employed group can dominate union policy (which seems plausible) the union is likely to push for higher wages, represented for instance by a move from B to C in Figure 2. In effect the utility mapping will be dependent upon the sequence of demand shifts. It can be shown what happens for the demand increase by altering the M-S objective function which is

$$\max L\big(U(W) - \overline{U}\big), \quad \text{subject to } R_L - W = 0,$$

where $\overline{U}$ = utility required to improve welfare for an unemployed union member.

The new objective function is

$$\max L_0\big(U(W)-\overline{U}\big) + \alpha(L-L_0)\big(U(W)-\overline{U}\big),$$

subject to $R_L - W = 0$,

where L_0 represents current employment.

The coefficient α $(0 \leqq \alpha \leqq 1)$ determines the relative influence of the unemployed union members on union policy. If α equals unity, the union operates a symmetric policy and the M-S results hold. If α is strictly less than unity, the new first-order condition (giving the tangency between L_D and the relevant indifference curve) will give a higher wage for isoelastic demand increases. This follows from the new first-order condition

$$BG_W/G(W) = \big[B + (1-\alpha)/\alpha\big] \times \big[U'(W)/\big(U(W)-\overline{U}\big)\big]$$

where $L_0 = G(W)$ and $L = BG(W)$.

Clearly when $\alpha < 1$ the indifference curves are flatter[2] as shown in Figure 2 and this ensures a higher wage in the new equilibrium.[3] The argument above relates to empirical or technical points which affect the M-S conclusions. However, the specification of union objectives raises a more fundamental issue in the theory of wage determination. In particular, wage theories which are based on the notion of wage targets by unions have been presented as an alternative theory to the orthodox (marginal productivity based) theories. For example, the Hicksian concept of real wage resistance (John Hicks, 1974) has encouraged an investigation of the relationship between money wage inflation and real wage aspirations. These studies are clearly competing with the Phillips curve relationship which ignores real wage targets. This comment is concerned with only one possible manifestation of this phenomenon where union leaders may in defense of their position conform to wage targets which are positively biased in periods of fluctuating demand. This is clearly important in the evaluation of the sticky wages hypothesis.

Finally, an aggregation problem exists when a one-union and one-employer model is used for macroeconomic predictions. The sticky wages result in the inefficient case occurs because of offsetting changes in the contract and equity loci. This depends on the independence of $\overline{W}$ (the minimum wage) and B (the level of demand). In general, we would expect an increase in the level of demand to result in an increase in all product prices. This implies that the net revenue of the particular firm may fall if its own product price rises by less than the general price index. What happens in these circumstances, and those where prices increase by more than the general price level, is even less obvious than with the original M-S assumptions where R_{LB} and R_B are always positive.[4] The fact that the outcome is uncertain is hardly convincing in this case, particularly as it results whether demand increases or decreases. In practice, it will depend on the parameter values. In these cases, we might still expect, in the face of uncertainty about the effect of demand increases on revenue and the marginal revenue product, that k, the fair share of wages, may still be a positive function of demand. This would still provide a basis for flexibility in the real wage level.

In conclusion, it is argued that the M-S results are based on a set of limited assumptions which restrict the generality of the model as an explanation for wage stickiness. Furthermore, if the macroeconomic criticisms are sustained, then even the results

[2]In the M-S model, $G_W = U'(W)/(U(W)-\overline{U})$.

[3]An additional constraint is that $L \geqq L_0$, and, if $\alpha = 0$, the relevant union objective is to maximize the wage subject to $L = L_0$. It is also possible that for some (small) values of α, the same wage may emerge. This does not affect the basic point that a weighting factor in favor of currently employed workers increases the chances of a higher wage.

[4]If R_{LB} and R_B are both negative (where the product price is rising by less than the general price level), the dW/dB outcome is still uncertain (see ft. 1). Note, however that the employment level will now fall as demand increases.

based on the same microeconomic assumptions are doubtful, because changes in demand will disturb the product prices of all firms and the implications for revenue changes are both unclear and perhaps unrealistic. A persistent influence on real wages which may be a more reliable factor is the union push for higher wages where demand increases. This wages-push may be due to the struggle to improve the wages share in revenue or be the result of union discrimination in favor of currently employed union members.

REFERENCES

Hicks, John R., *The Crisis in Keynesian Economics*, Oxford: Basil Blackwell, 1974.

McDonald, Ian M. and Solow, Robert M., "Wage Bargaining and Employment," *American Economic Review*, December 1981, *71*, 896–908.

Union Wage Policies: Reply

By IAN M. MCDONALD AND ROBERT M. SOLOW*

Paul Chapman and Malcolm Fisher make four separate remarks in their comment on our 1981 paper. One of their remarks points to a potentially important extension of our model; the others are more or less offhand asides. We comment in order.

1) If the union's bargaining power rises in demand upswings, wages are likely to be procyclical. Yes, of course. This can be seen in our most general bargaining model. If we employ the "generalized Nash solution" which maximizes a geometric mean of the parties' gains with weights depending on "bargaining power," then an increase in the weight attached to the union side will imply higher wages and more employment, *ceteris paribus*. The danger is that, in the absence of any direct measure of bargaining power, it becomes one of those self-sealing explanations. Do wages rise in an upswing? The union's bargaining power has increased. Do they not? Ah, this time it didn't.

2) In the monopoly-union case, the condition for wage stickiness is that the demand for labor shift isoelastically, and that is a fluke. A similar point can be made for the more complicated model of efficient bargains. In that model, the condition for the Nash solution to hold can be written

$$wu'(w)/(u(w)-u(\overline{w})) = (1+E_R)/(1-E_R).$$

where E_R is the elasticity of revenue with respect to employment.[1] If $\overline{w}$ is constant, then a condition for wage stickiness is that the revenue function shifts isoelastically (i.e., that E_R is constant). That is also a fluke. We are not at all certain that the union-monopoly model has realism on its side. In the United States at least, very few trade unions can name the wage unilaterally. It is true that unions may not wish to be seen trading off employment for higher wages. Our contention was that if the negotiated outcome is sticky wages and fluctuating employment, the parties may be content to negotiate a wage and allow the employer formally to determine employment, but be constrained by work rules that prevent employment from falling below the implicitly negotiated level. We did note that the literature contains arguments for procyclical demand elasticities (shift to relative luxuries as incomes rise) and countercyclical demand elasticities (less need to "economize" as incomes rise), and reckoned that perhaps there is not much cyclical variation in them.

3) If the union's objective function attaches more weight to the welfare of its currently employed members than its currently unemployed members, an asymmetry is introduced: the union is likely to favor wage gains over employment gains in an upswing because wage gains accrue to the currently employed, while gains in employment accrue to the currently unemployed. This is an interesting and important point. We have actually carried such a model quite far. The first draft of our paper included a section giving the results in some detail, but it was omitted from the published version to save space. The fundamental fact is that this sort of differential weighting introduces a traveling kink in the union's indifference curves, always at the current level of employment. In our model of efficient bargains, the kinked indifference curves introduce a vertical section into the contract curve (see Figure 1).

The equity locus is unaffected. The intersections of these new contract curves and

*Departments of Economics, University of Melbourne, Parkville, Victoria 3052 Australia, and Massachusetts Institute of Technology, Cambridge, MA 02139, respectively.

[1]The condition for the fair shares convention to hold is $wu'(w)[u(w)-u(\overline{w})]^{-1}=1-E_Rk^{-1}$, and has similar implications to the Nash solution.

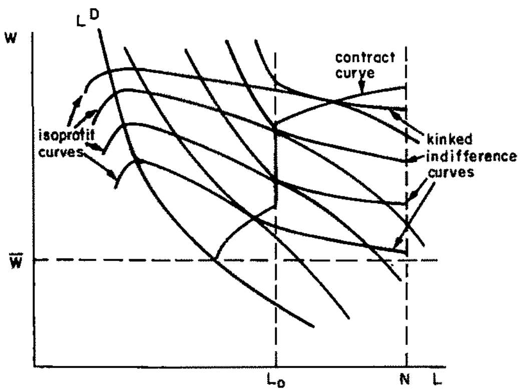


FIGURE 1

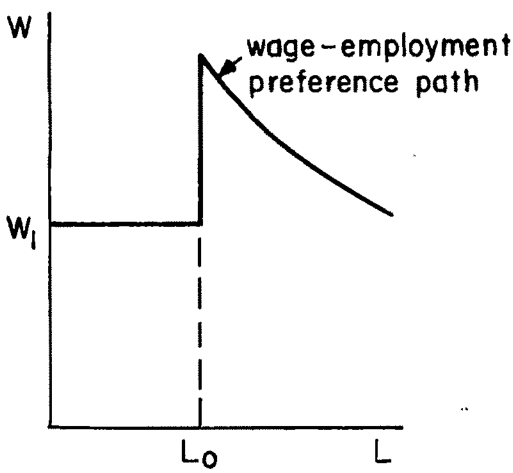


FIGURE 2

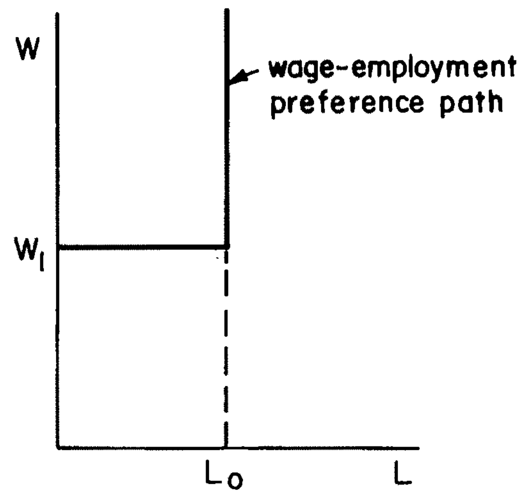


FIGURE 3

equity loci for various possible levels of labor demand greater or smaller than the initial level which produced $L = L_0$ traces out a wage-employment preference path that the trade union and firm will bargain for in the event of demand increases or decreases (compare A. M. Cartter, 1959, pp. 90–94). For neutral or isoelastic shifts of the firm's revenue function the union/firm wage-employment preference path will appear as in Figure 2. Declines in demand push wages to a floor—once on the floor, employment bears the brunt of adjustments to demand changes.

The effect of increases in demand is less clear cut and rather curious. Small demand increases raise wages alone, while large demand increases have a smaller positive effect on wages but do increase employment. Reducing the weight of the unemployed union members flattens the right-hand arm of the union indifference curves and increases the height of the vertical section of the wage-employment preference path. Eventually for a low enough weight of unemployed members the vertical section rises without limit and so the wage-employment preference path becomes a reverse L shape (see Figure 3). This asymmetry is extreme—demand increases fall entirely on wages, and demand reductions fall entirely on employment.

While the assumption of a lower weight for unemployed members appears realistic to us, the barrier it places on employment increases in this model seems too strong. Perhaps the assumption of fixed but different weights for employed and unemployed members of the pool is too simple. A better treatment would have to allow for seniority preferences and other such details.

4) At best, the model makes a case for wage stickiness in bargaining between a single employer or industry and its labor pool. Projection to the macroeconomic level may be

problematic because some firms' relative product prices will no doubt fall even in a general upswing. Very likely, there *is* a general equilibrium problem that our partial approach does not address. We stated as much explicitly. We would prefer to think of most markets in a fluctuating economy as imperfectly competitive with a number of prices adjusting gradually so that most firms could see their real demand curves shifting up together. In any case, we would be happy with the partial-equilibrium point.

REFERENCES

Cartter, A. M., *Theory of Wages and Employment*, Homewood: Richard D. Irwin, 1959.

Chapman, Paul G. and Fisher, Malcolm R., "Union Wage Policies: Comment," *American Economic Review*, September 1984, *74*, 755–58.

McDonald, Ian M. and Solow, Robert M., "Wage Bargaining and Employment," *American Economic Review*, December 1981, *71*, 896–908.

Government Debt: Comment

By WILLEM H. BUITER AND JEFFREY CARMICHAEL*

In a recent article in this *Review*, John Burbidge (1983) argues that our analyses of overlapping-generations models with intergenerational gifts and bequests (Buiter, 1979; 1980, and Carmichael, 1982) are logically faulty. On the basis of this assertion he goes on to argue that the real issue in the debate over debt neutrality is the nature of individual preferences. This comment argues that neither his criticism of our analyses nor his ultimate conclusion is valid.

I. Specification of the Utility Function

Burbidge's criticism of our treatment of intergenerational transfers centers on our specification of the individual's utility function. We assume that population grows at the rate n and that bequests (gifts) are shared equally among immediate descendants (ancestors). Logic then requires that $1+n$ enters the budget constraint in a perfectly obvious manner.

Nothing is implied by this about the relative merits of making my welfare an increasing function of the welfare of my *representative* descendant or ancestor (our specification), or of $(1+n)$ times the welfare of my representative descendant or $1/(1+n)$ times the welfare of my representative ancestor. It is an old and familiar issue in the optimal growth literature whether the utility of the representative individual should be optimized, or whether a "the more the merrier" approach should be adopted as in Burbidge. This is not an issue of logical consistency but of tastes. I might care about the welfare of my representative descendant raised to an exponent equal to the logarithm of the cube root of $1+n$. De gustibus non est disputandum.

*London School of Economics and Political Science, Houghton Street, London WC 2A 2AE, and Reserve Bank of Australia, 65 Martin Place, Box 3947, GPO 2001, Australia, respectively.

This, of course, is not the end of the story. The ultimate analytic results derived by ourselves and Burbidge depend not so much on the specification of the utility function per se as on the restrictions imposed on the parameters of the chosen utility functions. It is here that the real difference between ourselves and Burbidge lies.

Consider the case under contention, in which the individual cares only about his parents. Using Burbidge's notation, we can write the individual's utility, v_t, as a function of his own consumption in periods 1 and 2 (c^1 and c^2) and of the per capita utility of his parents, v_{t-1}. Under the "de gustibus..." principle, we can think of this latter relationship as some general function $F(\cdot)$ of v_{t-1}. In our earlier papers, we write the utility function in the form:

$$(1) \qquad v_t = u\left(c_t^1, c_t^2\right) + (1+\delta)v_{t-1},$$

where $(1+\delta)$ is the interpersonal discount factor applied to the *per capita* utility of the individual's parents. We will come back to the interpretation of this discount factor shortly.

Burbidge argues that, since each generation is $1+n$ times larger than the previous generation, parental welfare should be weighted by the factor $1/(1+n)$ in the descendant's utility function. Thus, he writes the utility function (his equation (3′)) as

$$(2) \qquad v_t = u\left(c_t^1, c_t^2\right) + \left(\frac{1+\rho}{1+n}\right)v_{t-1},$$

where ρ is again an interpersonal discount rate.

Equations (1) and (2) are a subset of the many acceptable utility functions in which v_t is a general function $F(\cdot)$ of v_{t-1}. What matters is the set of restrictions imposed on the derivative of this function, $\partial F/\partial v_{t-1}$. Clearly, any sensible restrictions on the in-

terpersonal discount factor in equation (2) will have counterpart restrictions on the discount factor in equation (1).[1] For this reason we will work with Burbidge's specification from here on.

Burbidge argues that $\rho > n$ in equation (2) is a necessary condition for sensible results. If he is correct, a steady state with gifts will have the property that $\rho = r > n$ (from the first-order condition for operative gifts). Such a steady state is necessarily undercapitalized, thereby ruling out the case in which bonds are an unambiguous addition to net wealth and, with it, the proof in Carmichael's 1982 paper that bonds can be neutral even when they are an unambiguous addition to net wealth. Burbidge's restriction, however, has some bizarre implications.

II. Restrictions on the Utility Function

There are two possible reasons for restricting the range of sensible values for ρ.[2] The first and most important source of restriction on ρ comes from the steady-state utility function, which from (2) is

$$v = \left(\frac{1+n}{n-\rho}\right) u(c^1, c^2).$$

A sensible model would have steady-state utility both bounded and of the same sign as $u(\cdot)$. Burbidge's condition, that $\rho > n$, however, does not yield a sensible steady state. If we assume a stationary consumption sequence $\{c_t^1, c_t^2\}$, utility defined by (2) with $\rho > n$, does not converge at all. Alternatively, if we assume that utility is stationary, the model has the peculiar characteristic that the steady-state utility function $v(\cdot)$ has the opposite properties to consumption utility $u(\cdot)$; for example, if $u(\cdot)$ is positive and increasing in c^1 and c^2, $v(\cdot)$ is negative and *decreasing* in c^1 and c^2.

Our specification of the utility function, which discounts parental welfare (i.e., $\rho < n$), is convergent for constant $\{c_t^1, c_t^2\}$ and the stationary utility level is, sensibly, increasing in c^1 and c^2. In terms of equation (1), (our specification) our restriction would be $\delta < 0$. The equivalent (sensible) restriction in Burbidge's formulation (equation (2)) is $\rho < n$.

The second potential source of restriction on ρ comes from consistency of the family's consumption plan; it is on this argument that Burbidge relies for his result. He asserts that consistency of the family's consumption plan requires that if the utility of heirs is discounted at ρ, the utility of parents must be "reverse discounted" at ρ. To establish this claim, he proposes a utility function (his equation (4)) which avoids the unpalatable implications of reverse discounting in his equation (3′) (equation (2) herein) by including concern for ancestors and descendants, but truncating direct and indirect concern about ancestral utility after the first generation:

$$\text{(3)} \quad v_t = \sum_{i=-1}^{\infty} \left(\frac{1+n}{1+\rho}\right)^i u\left(c_{t+i}^1, c_{t+i}^2\right) \quad \rho > 0.$$

While, under the de gustibus... principle, Burbidge is entitled to choose any utility function he likes, the asymmetry of equation (3) seems arbitrary. His justification for this truncation is unconvincing. Referring to equation (3) he argues, correctly, that "Since someone born at time t can affect only the utility level of his or her parents, v_{t-2} is exogenous at time t" (p. 223). However, the mere fact that I cannot affect something

[1] Indeed, if one had a preference for Burbidge's formulation, the discount factor in equation (1) could be interpreted as a weighted discount factor such that $(1+\delta) = (1+\rho)/(1+n)$. In fact, we see little to recommend Burbidge's formulation, since each individual always has one set of parents (not $1/(1+n)$) regardless of the population growth rate. However, since utility functions are a matter of taste, we see no reason to reject his specification out of hand.

[2] A third reason is intuitive. Since $1+\rho$ is an interpersonal discount rate it should conform to reasonable priors about interpersonal preferences. A value of ρ greater than zero implies that the individual "cares" more *per capita* for his parents than he does for himself. If, as in equation (2), the per capita utility of his parents is weighted by the factor $1/(1+n)$ he will care more *absolutely* (i.e., in total) for his parents than for himself if $\rho > n$. Under normal (selfish) preference theory we would expect the individual to apply a positive discount factor to the utility of others, suggesting the minimal restriction that $\rho < n$.

doesn't mean it cannot affect my utility. I may not be able to do anything about my toothache, or my father's toothache, or indeed a toothache suffered by my now deceased grandfather, but these events may well affect my utility. It is obviously correct that the optimal decision rule is independent of any additive term in the objective function which the decision maker takes as parametric. However, while exogenous to the individual, such terms are endogenous to the system as a whole and should not be dropped, as Burbidge proposes, when evaluating the long-run properties of the model.[3]

These objections notwithstanding, we can still evaluate Burbidge's claim about the discount rate using his formulation. Suppose, as Burbidge does, that individuals apply the discount rates $(1+\rho^-)$ to ancestors and $(1+\rho^+)$ to descendants. His "two-sided" utility function becomes

$$(4)\qquad v_t = \left(\frac{1+\rho^-}{1+n}\right)u_{t-1} + u_t + \sum_{i=1}^{\infty}\left(\frac{1+n}{1+\rho^+}\right)^i u_{t+i}.$$

So long as we assume that the individual takes the actions of others as given (i.e., a Nash concept of behavior), the relevant first-order conditions from maximizing (4) are

$$(5)\qquad \frac{\partial u_{t-1}/\partial c_{t-1}^2}{\partial u_t/\partial c_t^2} \le \frac{1+r_{t+1}}{1+\rho^-};$$

$$(6)\qquad \frac{\partial u_t/\partial c_t^2}{\partial u_{t+1}/\partial c_{t+1}^2} \ge \frac{1+r_{t+2}}{1+\rho^+}.$$

Burbidge argues that, in a steady state, an individual contemplating a gift to his parent will equate $(\partial u/\partial c_{t-1}^1)/(\partial u/\partial c_t^1)$ to $(1+r)/(1+\rho^-)$ while his parent, contemplating a gift to the individual will equate the same ratio of marginal utilities to $(1+r)/(1+\rho^+)$. Consistency of the family plan, he asserts, requires that these two ratios be equal; that is,[4] that $\rho^+ = \rho^-$. Since $\rho^+ > n$ is a necessary condition for (4) to converge to a sensible steady state, he also gets $\rho^- > n$.

Burbidge's conclusion, however, only holds if there are stationary equilibria in which both gifts and bequests are strictly positive. If this is the case, it should occur as a natural result of the model; it is not a condition that can sensibly be imposed. In this respect, the inequalities in (5) and (6) are crucial since they allow for corner solutions in gifts and/or bequests. If we impose the more plausible restrictions that $\rho^+ > n$ and $\rho^- < n$, the steady state has the characteristics that: if $r = \rho^- < n$, gifts are operative while bequests are zero; if $\rho^- < n < r = \rho^+$, bequests are operative, while gifts are zero; and if $\rho^- < r < \rho^+$ both gifts and bequests are at zero corner solutions. The actual steady-state configuration depends on tastes and technology.

Once the corner solution properties of intergenerational transfers are accounted for, it is no longer necessary for $\rho^- = \rho^+$, since it is no longer necessary for the two inequalities mentioned by Burbidge to hold with equality simultaneously. His is a special (we believe not very plausible) case.

It is worth emphasizing that, interpreted correctly, this model does allow for "dynamically inefficient" steady states, in which the marginal product of capital is below the population growth rate. This occurs if the stationary equilibrium has positive gifts from children to parents. The reasons that inefficient steady states can exist are that: (a) parental consumption and utility, even though it accrues earlier in time, is *discounted* by moderately selfish descendants; (b) agents act competitively in all markets; and (c) successive generations play a noncooperative game with each other. If binding commitments could be entered into between generations, a cooperative solution could be achieved that would rule out inefficient equilibria. That this does not occur is a funda-

[3]These issues and the nature of the "two-sided" utility function are considered in detail in Carmichael (1979).

[4]Our notation here varies slightly from Burbidge's. In his fn. 4, he uses $1/(1+\rho^-)$ instead of $(1+\rho^-)$ which, apart from being confusing, leads him to the incorrect (given his definition) restriction that $\rho^- = \rho^+$.

mental difference between overlapping generations models (even with interdependent preferences) and models with a single, infinitely lived individual.

III. Debt Neutrality: The Real Issue

On the basis of his claim that $\rho^- > n$, Burbidge argues that inefficient steady-state equilibria ($r < n$) are ruled out, thereby strengthening Robert Barro's (1974) debt-neutrality theorem. He argues that the essence of debt neutrality is therefore in the specification of the utility function. Apart from being incorrect in asserting the ρ^- must be less than n, we believe that Burbidge's conclusion misses the point of our contributions.

The objective of Carmichael (1982) was to establish that the literature was focussing attention on the wrong questions. As shown in that paper, whether or not individuals perceive government debt as a contribution to net wealth is irrelevant to the question of debt neutrality. Neutrality or nonneutrality is entirely a consequence of budgetary substitutability between debt and intergenerational transfers. The proof of that proposition consisted of showing that an undercapitalized equilibrium *could* exist and that, in this equilibrium, under suitable (restrictive) conditions government debt was still neutral, despite the fact that it was an unambiguous contribution to net wealth.[5] To refute this proposition, Burbidge would need to have shown that $\rho^- < n$ is *never* possible —which he has failed to do.

IV. Conclusion

Specification of utility functions is a matter of taste, not of logic. Analytic conclusions about steady-state behavior with intergenerational transfers, however, depend not so much on specification of the utility function as on restrictions imposed on the parameters of any given utility function; this is a matter of logic, not of tastes.

We see two main grounds on which to restrict interpersonal discount rates in Burbidge's (or any other) specification of the utility functions: the desirability of a sensible steady state; and consistency of the family consumption plan. Neither of these approaches leads to Burbidge's restriction that the utility of ancestors must be reverse discounted. Indeed they lead to the opposite conclusion.

Even allowing Burbidge's restriction as a possible (though not very plausible) case, his conclusion that debt neutrality revolves around specification of the utility function is still a red herring. The essence of debt neutrality is budgetary substitutability between intergenerational transfers and government debt.

[5] Burbidge does point out, correctly, that the suitable conditions are even more restrictive than suggested by Carmichael, in that his Propositions 1 and 2 should be amended to include the proviso that the *same* transfer mechanism must be operative both before and after the policy action.

REFERENCES

Barro, Robert J., "Are Government Bonds Net Wealth?," *Journal of Political Economy*, November-December 1974, *82*, 1095–117.

Buiter, Willem H., "Government Finance in an Overlapping-Generations Model with Gifts and Bequests," in G. M. von Furstenberg, ed., *Social Security Versus Private Saving*, Cambridge: Ballinger, 1979.

______, "'Crowding Out' of Private Capital Formation by Government Borrowing in the Presence of Intergenerational Gifts and Bequests," *Greek Economic Review*, August 1980, *2*, 111–42.

Burbidge, John B., "Government Debt in an Overlapping-Generations Model with Bequests and Gifts," *American Economic Review*, March 1983, *73*, 222–27.

Carmichael, J., "Economic Equilibrium and Steady-State Growth with Intergenerationally-Dependent Preferences," ERP Memo. No. 245, Princeton University, 1979.

______, "On Barro's Theorem of Debt Neutrality: The Irrelevance of Net Wealth," *American Economic Review*, March 1982, *72*, 202–13.

Government Debt: Reply

By John B. Burbidge*

In his debate with Martin Feldstein (1974, 1976) on the effects of government debt and Social Security, Robert Barro (1974, 1976) argued in effect that one had to ask how older people, some of whom had saved very little while working, survived in old age before Social Security existed. He postulated that individuals care about their parents and his model stressed the existence of a steady state with gifts from children to parents, in which government debt policy was irrelevant. If this steady state were overcapitalized (with the interest rate, r, less than the population growth rate, n), it would be easy for the government to use debt to raise the utility levels of both the current generations, and of all future generations. Most economists do not believe that real world economies are so blessed, and Barro was anxious to demonstrate that in his model's steady state with gifts, r exceeded n (1976, p. 345). All this notwithstanding, the centerpiece of recent papers by Willem Buiter (1979, 1980) and by Jeffrey Carmichael (1982), and of their comment on my paper, is that it is possible to have an overcapitalized steady state, with gifts, in which government debt is neutral. I shall confine my discussion to the following three points.

First, Buiter and Carmichael (hereafter, B-C) restate their position that if gifts are operative in the steady state, then $r < n$, but if bequests are operative, $r > n$, and that debt is neutral in either case. In their footnote 5, B-C observe that I was correct in asserting that Propositions 1 and 2 in Carmichael (p. 207) were logically faulty. These propositions require that the *same* intergenerational transfer mechanism be operative before and after the policy action. Apparently they accept my point (1983, pp. 225–26) that if a large quantity of government debt, or a generous Social Security program (certainly, neither the U.S. government debt nor the U.S. Social Security program is small!) were introduced into their steady state with gifts, gifts would be driven to zero and the interest rate would have to increase; government debt would *not* be neutral.[1] In my formulation of the Barro model, the introduction of *any* quantity of debt into a steady state (with either gifts or bequests initially) is neutral. What accounts for the difference? The difference is that I employed a utility function that is consistent in the sense that the first-order condition for a child contemplating a gift to his parents is the same as that for a parent contemplating bequests to his children.[2] B-C adopted a different utility function and consequently lost Barro's real debt-neutrality result.[3]

Furthermore, it is important to ask what "neutrality" means. Normally when one describes some governmental action as being "neutral" in its effects, it means that there is no rationale for taking that action. It is well known that overcapitalized steady states are not Pareto optimal in overlapping-generations models that assume individuals care only about their own consumption (see Peter Diamond, 1965, and David Gale, 1973). In particular, the introduction of government debt into these models would increase the welfare of everyone alive now and for all time to come. It turns out that this result

*McMaster University, Hamilton, Ontario L8S 4M4, Canada. I thank Peter Diamond, Peter Howitt and John Leach for helpful comments. The usual disclaimer applies.

[1] Likewise, a government program that made a sufficiently large transfer from the old to the young would not be neutral in their steady state with bequests, where r exceeds n.

[2] My utility function is consistent in Robert Strotz's sense; see my second point below.

[3] The logic of fn. 4 in my paper has eluded B-C. They insist that it relates to *steady states*, but it is clearly aimed at the *transition path* described above, i.e., from a steady state with gifts to a steady state with bequests. They also state that I assume that steady states exist "in which both gifts and bequests are strictly positive..." (p. 764), but, in fact, I explicitly rule out this possibility (p. 224).

carries over to B-C's steady state with gifts. In their setting, any rational government would introduce enough debt to drive gifts to zero, then introduce more debt, and make everyone better off in the process. So the B-C "debt-neutrality" result does not mean that debt policy is unimportant; on the contrary, their model provides a rationale for activist government policy.

Second, B-C object to my formulation of the individual's utility function in which the utility of the individual's parents is reverse discounted. As Robert Strotz demonstrated long ago, consistent behavior requires exponential discounting (see 1956, pp. 174–75). If the individual cares about his ancestors and his descendants, then his utility function must be of the form $v_t = \sum_{i=-N}^{\infty} \beta^i u_{t+i}$, where $N \geq 1$. Since the addition of constants to a utility function does not affect behavior, and for someone born at time t, each element of $\{u_t\}_{t=-\infty}^{-2}$ is a constant, I used the utility function $\sum_{i=-1}^{\infty} \beta^i u_{t+i}$, where $\beta = (1+n)/(1+\rho)$ (equation (4) in my paper). This is obviously well-defined for values of β less than unity or, equivalently, ρ greater than n. To repeat, my utility function is consistent in Strotz's sense.

Third, B-C argue that "[t]he essence of debt neutrality is budgetary substitutability between intergenerational transfers and government debt" (p. 765), rather than the specification of the individual's utility function. Suppose individuals in the Diamond model were permitted to make gifts to their parents or bequests to their children. Since these agents are selfish by assumption, there would be neither gifts nor bequests and the model's results would be totally unaltered. That government debt is neutral in Barro's model but not in Diamond's originates primarily from differences in the specification of the individual's utility function. Buiter and Carmichael are incorrect in arguing otherwise.

To summarize, Barro's intuition is correct —a formulation of his model, which yields his result that *any* quantity of debt is neutral, cannot have overcapitalized steady states. Barro's model is essentially the optimal growth model (see, for example, Diamond, 1973). In steady state this model exhibits the "modified Golden Rule," and as a consequence, the interest rate must exceed the population growth rate.

REFERENCES

Barro, Robert J., "Are Government Bonds Net Wealth?," *Journal of Political Economy*, November-December 1974, *82*, 1095–117.

______, "Reply to Feldstein and Buchanan," *Journal of Political Economy*, April 1976, *84*, 343–49.

Buiter, Willem, "Government Finance in an Overlapping-Generations Model with Gifts and Bequests," in George M. von Furstenberg, ed., *Social Security Versus Private Saving*, Cambridge: Ballinger, 1979.

______, "'Crowding Out' of Private Capital Formation by Government Borrowing in the Presence of Intergenerational Gifts and Bequests," *Greek Economic Review*, August 1980, *2*, 111–42.

______ **and Carmichael, Jeffrey**, "Government Debt: Comment," *American Economic Review*, September 1984, *74*, 762–65.

Burbidge, John B., "Government Debt in an Overlapping-Generations Model with Bequests and Gifts," *American Economic Review*, March 1983, *73*, 222–27.

Carmichael, Jeffrey, "On Barro's Theorem of Debt Neutrality: The Irrelevance of Net Wealth," *American Economic Review*, March 1982, *72*, 202–13.

Diamond, Peter, "National Debt in a Neoclassical Growth Model," *American Economic Review*, December 1965, *55*, 1126–50.

______, "Taxation and Public Production in a Growth Setting," in J. A. Mirrlees and N. H. Stern, eds., *Models of Economic Growth*, New York: Wiley & Sons, 1973.

Feldstein, Martin, "Social Security, Induced Retirement and Aggregate Capital Accumulation," *Journal of Political Economy*, September-October, 1974, *82*, 905–26.

______, "Perceived Wealth in Bonds and Social Security: A Comment," *Journal of Political Economy*, April 1976, *84*, 331–36.

Gale, David, "Pure Exchange Equilibrium of Dynamic Economic Models," *Journal of Economic Theory*, February 1973, *6*, 12–36.

Strotz, Robert, "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies*, No. 3, 1956, *23*, 165–80.

Accumulation of Property by Southern Blacks before World War I: Comment and Further Evidence

By ROBERT A. MARGO*

The pace and pattern of wealth accumulation among southern blacks from emancipation to World War I is of central importance to the historical evolution of black-white income differences. In a recent paper in this *Review* (1982), Robert Higgs uses Georgia tax assessment records to show that blacks accumulated property more rapidly than whites during this period, subject to short-run fluctuations in cotton income. Cross-sectional regressions reveal that black wealth varied inversely with illiteracy, the price of land, and the tenancy rate; positively with cotton cultivation, the presence of plantation agriculture, and population density; but was insignificantly affected by racial composition. As Higgs points out, "these findings call into question the traditional interpretation of the role played by the plantation-cotton-black-belt complex" (p. 735), and underscore the long-run effects of illiteracy and educational discrimination on black economic progress.

This comment extends Higgs' findings in two directions. First, race-specific data on assessed wealth are examined for five additional states. Analysis of these data reveals that the Georgia time-series pattern of rapid growth of black assessed wealth, absolutely and relative to white assessed wealth, is a more general finding (with one exception), but that cross-sectional regressions of black assessed wealth yield only limited conclusions once interstate comparisons are made. Second, because the data are assessed and not true wealth, the sensitivity of the results to cross-sectional and temporal variation in assessment ratios is an important issue. Evidence for one state demonstrates that failure to control for intrastate variation in assessment ratios may bias the cross-sectional results, and that blacks faced higher average assessment ratios than whites. Furthermore, preliminary upper bound calculations suggest that adjusting for changes in relative (black-white) assessment ratios may significantly reduce the relative rate of growth of black wealth, although the substantive conclusion—that blacks accumulated wealth more rapidly than whites—still holds.

Early in his paper, Higgs refers to Georgia as "the only state with a long and highly disaggregated record of race-specific property holdings" (p. 726). Beginning in the late nineteenth century, however, and continuing beyond 1915 (the final year of Higgs' study), race-specific figures on assessed wealth similar to the Georgia data are available for Louisiana, North Carolina, and Virginia.[1] In addition, the State of Kentucky Auditor *Report*(s) contain race-specific wealth data from 1866 to 1885, and race-specific figures for Arkansas from 1895 to 1911 can be computed from the data on total assessed wealth and the black share of property tax payments given in its auditor's reports.

Figure 1 (analogous to Higgs' Figure 1) charts the total assessed value of property (in 1910–14 dollars) held by blacks in each state for the years data are available (or can be

*Assistant Professor of Economics, University of Pennsylvania, Philadelphia, PA 19104 and Faculty Research Fellow, National Bureau of Economic Research. I am grateful to Stanley Engerman, Claudia Goldin, Robert Higgs, Paul Menchik, Kenneth Sokoloff, Richard Sylla, Paul Taubman, and Gavin Wright for helpful comments. An earlier version of this paper was issued as NBER Working Paper No. 1200. This research was supported by the University of Pennsylvania. All errors are my own.

[1]See Tipton Ray Snavely (1916) and Samuel Bitting (1915) for two early twentieth-century studies making use of the Virginia data. J. Morgan Kousser (1980a,b) uses the North Carolina and Kentucky data to study the effects of changing black political power on the racial division of school tax burdens and public school expenditures in the late nineteenth and early twentieth century; see my 1982a article for a similar study based on the Louisiana data.

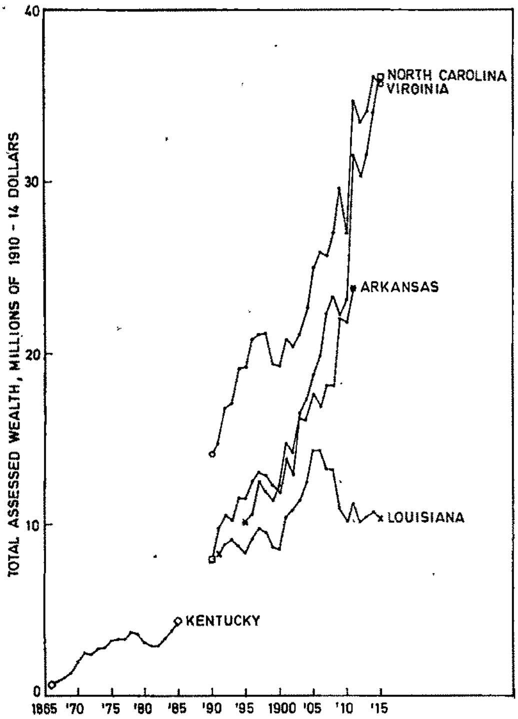


FIGURE 1. TOTAL ASSESSED VALUE OF BLACK WEALTH: NORTH CAROLINA; VIRGINIA; ARKANSAS; LOUISIANA; KENTUCKY

estimated) to 1915, and Table 1 (analogous to Higgs' Table 1) gives race-specific per capita estimates for the census and other years in this time interval.[2] A defect of the Louisiana records is the absence of race-specific data for New Orleans; the figures in brackets adjust for this omission, based on the assumption that the black-white ratio of per capita wealth in New Orleans was identical to the ratio prevailing elsewhere in the state.[3] The Arkansas estimates should also be viewed with some caution as they may be sensitive to the time-series pattern of property tax delinquencies, although the fluctuations are similar in magnitude to those observed in the other states.

The data reveal that, despite socioeconomic and political differences across regions, blacks in every state experienced rapid growth in total assessed wealth, and with the exception of Louisiana, in per capita terms, absolutely and relative to white assessed wealth. Other than a trough in the late 1870's, the Kentucky data suggest that substantial progress was made in the immediate post-emancipation period. In contrast to Georgia and Louisiana, blacks in Arkansas, North Carolina, and Virginia continued to augment their per capita wealth in the 1890's, despite a common downturn late in the decade. With the striking exception of Louisiana, per capita black wealth increased rapidly after 1900, and the black-white ratio of per capita wealth continued to rise. The boll weevil wreaked havoc on Louisiana's cotton crop very early in the century: between 1908 and 1915, average annual cotton income (in 1910–14 dollars) fell by 49.2 percent compared to the previous eight years. The significant positive association that Higgs found between cotton income and black wealth in Georgia suggests that a similar explanation might apply to Louisiana. This hypothesis is confirmed by the following regression (absolute value of t-statistics are shown in parentheses):

$$LBW = \underset{(3.25)}{-0.73} + \underset{(5.10)}{0.59}\, LBW(-1) + \underset{(3.65)}{0.07}\, CY(-1) + \underset{(1.78)}{0.08}\, YDUM$$

$$N = 24, \quad \bar{R}^2 = 0.80;$$

where LBW is the log of total black wealth in Louisiana (in millions, 1910–14 dollars), $LBW(-1)$ is LBW once lagged, $CY(-1)$ is total cotton income in Louisiana (in millions, 1910–14 dollars) once lagged, and $YDUM$ is a dummy variable for observations from

[2] Following Higgs, the George Warren-Frank Pearson wholesale price index (1933, pp. 12–13) is the deflator.

[3] An alternative procedure is to use the Louisiana cross-sectional regression (see Table 2) to predict per capita black wealth in New Orleans in 1910, and assume that the ratio (per capita black wealth in New Orleans/per capita black wealth, all other parishes) remained constant over the period. Although the levels are sensitive to the method employed, the growth rate of black wealth is not, and the conclusions reached in the text are unaffected.

TABLE 1—PER CAPITA ASSESSED WEALTH, 1870–1910[a]

State	1870	1880	1885	1890	1895	1900	1910	$r_g \times 100$
Arkansas								
White					267.53	247.80	307.39	0.9
Black					29.96	33.15	49.14	3.3
Ratio					8.9	7.5	6.3	−2.4
Kentucky								
White	262.69	252.28	306.95					1.0
Black	7.38	11.51	16.07					5.2
Ratio	35.6	21.9	19.1					−4.2
Louisiana								
White				294.67		300.82	401.41	1.6
				[472.46]		[435.99]	[595.12]	[1.2]
Black				16.46		14.86	16.31	−0.04
				[20.13]		[17.83]	[20.03]	[−0.02]
Ratio				17.9		20.2	24.6	1.64
				[23.5]		[24.5]	[29.7]	[1.22]
North Carolina								
White				241.64		241.70	304.27	1.2
Black				14.07		19.01	33.12	4.3
Ratio				17.2		12.7	9.2	−3.1
Virginia								
White				429.55		417.44	385.65	−0.5
Black				22.23		29.17	40.17	3.0
Ratio				19.3		14.3	9.6	−3.5

Notes and Sources: r_g: Average annual rate of growth from beginning to end date (for example, Louisiana, 1891–1915). Kentucky: Auditor of Public Accounts..., *Report*, 1866–85. Population figure for 1880 was interpolated from 1860 and 1880 census totals. Louisiana: Auditor of Public Accounts..., *Biennial Report*, 1891–1915. The "1890" estimate is for 1891; population figures for 1891 were interpolated from 1890 and 1900 census totals. Wealth estimates for 1894 and 1898 were interpolated from adjacent years; data are missing in the auditors' reports for these years. "Adjusted" estimates (in brackets) include imputed wealth for New Orleans parish (see fn. 3). North Carolina: Auditor of the State..., *Annual Report*, 1890–99; North Carolina Corporation Commission..., *Report*, 1900–15. Estimates for 1890, 1892–94 were constructed by multiplying the black share of property taxes and total assessed wealth. Virginia: Auditor of Public Accounts..., *Annual Report*, 1890–1915. The 1890 wealth estimated from data on personal property and 1891 share of personal property in total assessed wealth (blacks, 0.35; whites, 0.32); 1892, 1895 data on personal property estimated by multiplying black share of personal property taxes and assessed value of personal property. Arkansas: Auditor..., *Biennial Report*, 1895–1911. Wealth estimated by multiplying black share of property taxes and total assessed wealth. Estimates for 1898, 1900, 1902, 1904 are interpolated from adjacent years; data missing in auditor's report. Population: All population data are from U.S. Bureau of the Census (1918, pp. 43–44).
[a]Shown in 1910–14 dollars.

1908–15.[4] If cotton income is excluded, the coefficient of *YDUM* is -0.05, ($t = 1.25$).

In summary, the evidence from these additional states (with the exception of Louisiana) would tend to support one of Higgs' major conclusions—that southern blacks rapidly

[4]Cotton income is measured as gross producer revenue (millions, 1910–14 dollars), derived from December 1 price data in U.S. Department of Agriculture (USDA) (1927, p. 207) and quantity data in USDA (1951, p. 79). The regression specification is similar to Higgs', but with cotton income lagged once to preserve degrees of freedom. For further evidence on the boll weevil's impact on early twentieth-century cotton production, see Roger Ransom and Richard Sutch (1977, Table 9.2, p. 175). *LBW* excludes estimated wealth for New Orleans (see fn. 3); the qualitative results are unchanged if it is included. Cotton income was also significantly related to black wealth in Arkansas ($t = 2.29$) and North Carolina ($t = 2.25$). These additional regressions are available from the author on request.

TABLE 2—REGRESSIONS OF PER CAPITA WEALTH AND WEALTH PER ADULT MALE: 1910

	Arkansas	Georgia		Louisiana	North Carolina		Virginia			
	WPC (1)	*WPC* (2)	*WPAM* (3)	*WPC* (4)	*WPC* (5)	*WPAM* (6)	*WPC* (7)	*WPC* (8)	*WPAM* (9)	*WPAM* (10)
Constant	35.06	50.00	204.15	4.90	54.49	265.42	54.88	41.03	179.94	119.91
	(1.33)	(11.74)	(10.52)	(0.66)	(8.65)	(9.26)	(5.99)	(4.09)	(4.28)	(2.53)
PLANTATION	6.83	5.16	28.12	2.56	−5.50	−26.23				
	(0.44)	(2.92)	(3.44)	(1.00)	(2.17)	(2.27)				
$ACRE	0.95	−0.56	−2.70	−0.02	−0.46	−2.60	−0.007	−0.007	−0.03	−0.03
	(1.33)	(3.65)	(3.89)	(0.15)	(2.74)	(3.44)	(4.82)	(4.65)	(4.73)	(4.51)
BP10ILLIT	0.32	−0.29	−1.45	−0.04	−0.59	−2.88	−0.33	−0.43	0.14	−0.16
	(0.47)	(2.71)	(2.93)	(0.03)	(3.11)	(3.33)	(1.24)	(1.65)	(0.10)	(0.13)
POP/SQMI	−0.51	28.99	144.01	98.21	3.51	2.87	1.10	0.15	2.62	−1.09
	(1.72)	(2.43)	(2.69)	(1.72)	(1.32)	(0.26)	(2.40)	(0.27)	(1.30)	(0.44)
BPOP/POP	−52.01	−4.75	−24.54	−18.40	25.07	107.62	17.11	1.52	74.86	14.50
	(1.62)	(0.94)	(1.06)	(2.62)	(3.87)	(3.68)	(1.75)	(0.14)	(1.68)	(0.29)
COT/IMPAC	32.30	25.15	106.30	−1.14	11.18	57.23				
	(0.54)	(2.94)	(2.79)	(0.06)	(0.87)	(0.99)				
BTEN/BM21	−24.20	−30.72	−76.88	18.53	−6.60	17.83	−68.64	−64.40	−235.77	−219.74
	(0.53)	(5.77)	(3.14)	(1.81)	(0.84)	(0.50)	(4.82)	(4.23)	(3.15)	(3.00)
ASR								56.57		228.36
								(2.88)		(2.52)
BM21/BPOP	100.88			54.32						
	(4.42)			(5.44)						
R^2	0.40	0.39	0.25	0.37	0.36	0.38	0.42	0.46	0.24	0.28
N	44	130	130	57	81	81	98	98	98	98

Notes: Absolute value of t-statistics are shown in parentheses. *WPC*: per capita wealth; *WPAM*: wealth per adult male; *ASR*: weighted average of assessment ratios on real and personal wealth; *BM21/BPOP*: black adult males/total black population: *PLANTATION* = 1 if plantation county, 0 otherwise; *$ACRE*: per acre census value of farmland; *BP10ILLIT*: percentage illiterate of black population aged 10 or more; *POP/SQMI*: total population (in 1,000s) per square mile; *BPOP/POP*: percent black; *COT/IMPAC*: cotton acreage as a proportion of all improved acreage; *BTEN/BM21*: black farm tenants as a proportion of black males aged 21 or more. *WPC*, *WPAM*, and *$ACRE* are given in 1910–14 dollars (Warren-Pearson wholesale price index is the deflator).

accumulated assessed wealth during the postbellum period, and succeeded in closing the gap with white assessed wealth in relative terms.

Table 2 reports county-level weighted least squares regressions of black assessed wealth in 1910 (per adult male and/or per capita) for Arkansas, Georgia, Louisiana, North Carolina, and Virginia. The regression specification is identical to Higgs' with four exceptions.[5] First, the plantation county dummy (*PLANTATION*) and the share of improved acreage devoted to cotton (*COT/IMPAC*) were excluded from the Virginia regressions because of an insufficient number of plantation counties (two), and because cotton was a minor crop in Virginia. Second, including the proportion of black adult males in the total black population (*BM21/BPOP*) as a control for differences in the age-sex composition across counties considerably improved the fits of the Arkansas and Louisiana regressions.[6] Third, an estimate of the average assessment ratio on real estate and personal property (*ASR*) is added to two of the Virginia regressions.[7] The logic here is sim-

[5]The regressions also follow Higgs by averaging the dependent variables over 1909–11 and by excluding counties with black populations less than 1,000.

[6]If this variable is excluded, the F-ratios for both regressions fall short of the critical value at the 5 percent level of significance. Higgs controlled for differences in the age-sex composition across Georgia counties by estimating separate regressions of per capita wealth and wealth per adult male; this procedure was successful in the case of North Carolina and Virginia.

[7]Caution should be exercised in interpreting the Arkansas and Louisiana results since the cross-sectional reliability of the data for these states cannot be investigated by Higgs' method (pp. 727–28) or by controlling directly for assessment ratio variation. See the text.

ple: assessed wealth is the product of the assessment ratio and the true value of wealth, and the regression coefficients may confound the effects of the independent variables on these two components. This problem is potentially serious in cross-sectional analyses of assessed wealth, because assessment ratios varied greatly from county to county. By controlling for the assessment ratio in the regression, biases can be identified by examining the effects on the other regression coefficients.

Fourth, Higgs argued that the cross-sectional error terms were heteroscedastic, due to variation across counties in black population sizes. To address this problem, Higgs first weighted the Georgia data by the square root of the relevant population size (this procedure has also been followed for the other states), but incorrectly included both a constant term *and* the *GLS* weight in his regressions (see Table 3, p. 735). Only the latter is appropriate, and its coefficient is the estimate of the constant term (G. S. Madalla, 1977, p. 268; Potluri Rao and Roger Miller, 1971, p. 121). Corrected estimates for Georgia appear in Table 2. Fortunately, the corrected coefficients closely resemble Higgs' in sign and statistical significance, with the exception of the black population share, which reverses sign but remains statistically insignificant.[8] The corrected $\bar{R}^2$s, however, are much lower than those reported by Higgs.

Inspection of Table 2 reveals that the signs and magnitudes of the coefficients and their associated significance levels vary widely across states, and few, if any, generalizations are possible. If attention is restricted to the results for North Carolina, Georgia, and Virginia, however, some limited conclusions may be drawn.[9] Specifically, high land values were significantly associated with lower levels of black wealth, as were illiteracy and tenancy (in two of three states), and there is no evidence that an increase in the proportion black or the cotton share significantly *reduced* black wealth.

Equations (8) and (10) in Table 2 demonstrate that controlling for the assessment ratio eliminates the positive effects of population density and racial composition on black wealth in Virginia, and increases the negative effect of illiteracy. This suggests that assessment ratios in Virginia were higher in heavily black counties, urban areas, and counties where the black illiteracy rate was high. Two of these three hypotheses are confirmed in the following regression (absolute value of t-statistics are shown in parentheses):

$$\underset{(7.49)}{ASR = 0.028} + \underset{(4.63)}{0.19}\, BPOP/POP$$

$$+ \underset{(7.33)}{0.002}\, POP/SQMI + \underset{(1.31)}{0.0015}\, BP10ILLIT$$

$$N = 101, \quad \bar{R}^2 = 0.42.$$

Higher assessment ratios in cities are not very surprising, but the significant positive

[8]The assessment ratios were estimated by the Virginia Tax Commission, whose function was "to consider and report on all questions pertaining to the assessment and collection of taxes," and are given in Joint Committee on Tax Revision, Virginia, *Report* (1914, pp. 11–12, 251). The assessment ratios on real estate were derived by comparing the sales prices and assessed values of 20,694 transactions occurring between February 1, 1912 and January 31, 1913. The assessment ratios on personal property refer to farm machinery and livestock, and were derived by comparing the assessed values with 1910 census data on a per unit basis. Because Virginia law provided for a complete reassessment of all property every five years, had the committee applied the same procedures to data collected before the last reassessment (in 1910), a different cross-sectional pattern in assessment ratios may have existed. Ideally, assessment ratios for each year (1909–11) should be averaged into a single measure, but the data needed for such a calculation are unavailable. Furthermore, while the committee collected data on the race of the taxpayer, the published assessment ratios (used in the Virginia regressions) are not race specific. In the regressions, ASR is set equal to $a_r ASR_r + (1-a_r) ASR_p$, where a_r is the share of real estate in total assessed value, ASR_r is the assessment ratio on real estate, and ASR_p is the assessment ratio on personal property. The regression results are unaffected if ASR_r is substituted for ASR.

[9]Four Georgia counties with black population majorities (Camden, Glynn, Liberty, and McIntosh) specialized in rice production, grew virtually no cotton, had relatively high levels of per capita black wealth and low rates of tenancy, and were geographically distinct from the rest of the black belt. If a dummy variable distinguishing these counties ($RICE$) is added to the Georgia regressions, the coefficient of percent black becomes a larger negative number, and is significant at the 15 percent level in equation (3).

TABLE 3—RACE DIFFERENCES IN ASSESSMENT RATIOS: VIRGINIA, 1912

	Counties		Cities	
Sale Price	*N*	*ASR*	*N*	*ASR*
Less than $500				
White	6047	44.9	772	58.5
Black	1636	54.5	149	66.2
Ratio		0.82		0.88
$500–999				
White	2688	38.1	730	57.8
Black	277	48.4	105	61.2
Ratio		0.79		0.94
$1,000–2,499				
White	3083	36.4	1217	56.5
Black	136	38.3	63	54.4
Ratio		0.95		1.04
$2,500+				
White	2461	30.5	1380	51.7
Black	34	32.3	15	57.7
Ratio		0.94		0.90
Total (All Transactions)				
White	14279	33.1	4099	52.9
Black	2083	45.3	332	58.7
Ratio		0.73		0.90

Notes: N = Number of real estate transactions; ASR = average assessment ratio (Total sales/Total assessed value). *Source:* Snavely, pp. 75–76.

association with percent black seems peculiar. An early study by Tipton Ray Snavely, based on data collected by the Virginia Tax Commission (see fn. 8), provides some clues. Snavely discovered that assessment ratios on Virginia real estate varied inversely with the sales price of the property; holding constant sales price, black property was assessed at higher ratios than white property, except in one valuation class (see Table 3). Thus the higher assessment ratios in the Virginia black belt may be partly due to racial composition, although verifying this conjecture would require that Snavely's data be racially disaggregated at the county level, which cannot be done from published sources. The overall positive association between race and the assessment ratio may reflect the political nature of the assessment process (Virginia blacks were disenfranchised by 1910), and the relative ease of determining the market value of small holdings of land. "The true value... of small and little-improved holdings," wrote the authors of the Virginia Tax Commission's *Report*, "is much easier to ascertain than is that of the rich man's estate. The poor man, furthermore, usually has for his protection little influence, either personal or political. Finally, the poor man is ignorant of the means of correcting an unfair assessment or finds he cannot afford it" (1914, p. 10).

A specific implication of this discussion is that ignoring the intrastate variation in assessment practices would lead to the erroneous inference that Virginia blacks in black-belt counties or urban areas were more successful in accumulating property than blacks elsewhere in the state, all other things held constant. Whether similar cross-sectional biases are present in the other states is unclear. Higgs tested the cross-sectional reliability of the Georgia data by examining the fit (and outliers) of a regression of the per acre census value (in 1910) of farm real estate reported by black farm owners on the per acre assessed value of real estate (in 1910) listed by black taxpayers in 1910. By this method the North Carolina data would be judged reliable (this test cannot be performed for

Arkansas or Louisiana), as the fit of the regression was similar to the Georgia case.[10] Alternatively, Higgs' test implicitly defines the assessment ratio on black farm real estate to be the ratio of the per acre assessed value and the per acre census value. Adding this variable to the North Carolina and Georgia regressions had no significant effects on the other coefficients, and the own coefficient was always insignificant.[11] Errors in measuring the assessment ratio in this manner, however, may be partly responsible for these results. In Georgia, for example, the acreage figures reported by black taxpayers exceed the census farm acreage figures in all but five counties, especially in heavily black counties and urban areas. Thus some of the land reported to the Georgia tax authorities may not have been in agricultural use, particularly in urban counties, and assigning to it the per acre value indicated by the census farm data may be incorrect.[12]

The major implication of these findings, however, is that the relative level of black wealth in 1910 is clearly overstated by the tax data. Furthermore, given that state and local government discrimination against southern blacks evidently increased around the turn of the century (see, for example, C. Vann Woodward, 1955; J. Morgan Kousser, 1974, 1980a; my 1982b dissertation), the possibility that the relative rate of growth of black wealth is biased upwards by increasingly discriminatory assessment ratios cannot be discounted.[13] While a definitive answer to this question is beyond the scope of this comment, some preliminary calculations suggest that the bias may be significant, although not large enough to affect the substantive time-series conclusions. To illustrate this point in the Virginia case, assume that the relative assessment ratio on all taxable wealth was one in 1890 (surely a lower bound), and at the average value implied by Snavely's figures for real estate (1.22) in 1910. Under this assumption, increasingly discriminatory assessment ratios could account for as much as 28.6 percent of the relative growth of black per capita wealth in Virginia over these two decades.[14] While this figure is not trivial, and indicates the need to be cautious before assuming racial stability in assessment ratios over time, the magnitude of the bias is not sufficient to reverse the

[10]The correlation coefficient (r) between the tax and census values was 0.55 in North Carolina, compared to 0.62 in Georgia (Higgs, p. 727). Eliminating outliers (following the same procedures outlined by Higgs) from the North Carolina regressions had no effect on the regression results.

[11]On the other hand, eliminating counties from the Georgia regressions that Higgs identified as having much lower or higher than average assessment ratios on farm real estate yielded significant *negative* coefficients on population density, and insignificant coefficients on the cotton share, land value, and the plantation dummy (except equation (3)). Furthermore, the coefficients of the cotton share and population density are extremely sensitive to a single observation, Chatham, which contains the city of Savannah. If this county is deleted, both coefficients are sharply reduced in magnitude and become statistically insignificant. These additional regressions are available from the author on request.

[12]The mean value of the ratio of the census and tax acreage figures across all Georgia counties in 1910 was 0.78. In the five urban counties that Higgs identified plus Chatham County the mean was 0.65. The mean was 0.63 in heavily black counties (75 percent black or higher). W. E. B. Du Bois asserted that the discrepancy between the tax and census acreage figures could be accounted for by "land sublet by negro owners to tenants" (1906, p. 526). This seems implausible in urban counties, and does not explain why the number of acres listed by white taxpayers exceeds the number of acres owned by whites as reported in the 1910 census of agriculture (defined as the total acreage in farms less acreage reported by black farm owners). Also, it is worth noting that the Georgia auditors' reports never explicitly state that the acreage figures refer only to farm land. This explanation differs from Higgs, who questioned the accuracy of the census data, arguing that farmers in urban counties valued their land higher for tax purposes ("taking into account the near-term prospects of conversion to more valuable nonagricultural uses," p. 728) than in reporting to the census, or that some farms fell within city limits, thereby facing higher assessment ratios.

[13]Redistribution of the black population towards areas with above average assessment ratios could have raised the average assessment ratio on black property, even if discrimination did not change. Given that the percentage of blacks residing in the black belt declined between 1880 and 1910, and that the rate at which whites urbanized was greater than blacks, geographic redistribution probably *lowered* the relative assessment ratio, *ceteris paribus*.

[14]This figure would be biased upwards if, as Snavely argued (p. 90), personal property was assessed in a nondiscriminatory manner.

principal conclusion that blacks outpaced whites in accumulating wealth.[15]

This comment has presented additional evidence on wealth accumulation by southern blacks in the period before World War I. Analysis of these data indicates that blacks generally accumulated property at a more rapid pace than whites, but that only limited conclusions may be drawn from cross-sectional regressions of black assessed wealth. Evidence for one state indicates that failure to control for within-state variation in assessment practices may bias the inferences drawn from the cross-sectional regressions, and that black property was assessed closer to market value than white property. Finally, accounting for changes in relative assessment ratios over time may significantly reduce the relative rate of growth of black wealth, although not enough to reverse the substantive time-series results. In light of these findings, a useful extension of this work would be to estimate cross-sectional regressions of black wealth for different years, and to further analyze the determinants and implications of intrastate and race differences in assessment ratios.

[15] Preliminary calculations suggest that the bias in the Georgia case could be much larger. Despite the reservations noted in the text and in fn. 9, Higgs' procedure can be used to calculate the relative assessment on farm real estate in Georgia for 1910; the result is 1.46. If the relative assessment ratio were one in 1880, the relative growth rate of black wealth in Georgia from 1880 to 1910 would be overstated by 48 percent (1.4 percent compared to 2.7 percent), but again the substantive conclusions remain the same.

REFERENCES

Bitting, Samuel T., *Rural Landownership Among the Negroes of Virginia, With Special Reference to Albermarle County*, Charlottesville: University of Virginia Press, 1915.

Du Bois, W. E. B., "The Negro Farmer," in *Twelfth Census of the United States: 1900, Special Reports, Supplementary Analysis, and Derivative Tables*, Washington: USGPO, 1906.

Higgs, Robert, "Accumulation of Property by Southern Blacks before World War I," *American Economic Review*, September 1982, *72*, 725–37.

Kousser, J. Morgan, *The Shaping of Southern Politics: Suffrage Restriction and the Establishment of the One Party South, 1880–1910*, New Haven: Yale University Press, 1974.

______, (1980a) "Progressivism—For Middle Class Whites Only: North Carolina Education, 1880–1910," *Journal of Southern History*, May 1980, *46*, 169–94.

______, (1980b) "Making Separate Equal: Integration of Black and White School Funds in Kentucky," *Journal of Interdisciplinary History*, Winter 1980, *10*, 399–428.

Madalla, G. S., *Econometrics*, New York: McGraw-Hill, 1977.

Margo, Robert A., (1982a) "Race Differences in Public School Expenditures: Disfranchisement and School Finance in Louisiana, 1890–1910," *Social Science History*, Winter 1982, *6*, 9–33.

______, (1982b) "Disfranchisement, School Finance, and the Economics of Segregated Schools in the U.S. South, 1890–1910," unpublished doctoral dissertation, Harvard University, 1982.

Ransom, Roger L. and Sutch, Richard, *One Kind of Freedom: The Economic Consequences of Emancipation*, New York: Oxford University Press, 1977.

Rao, Potluri and Miller, Roger L., *Applied Econometrics*, Belmont: Woodsworth Publishing, 1971.

Snavely, Tipton Ray, *The Taxation of Negroes in Virginia*, Charlottesville: University of Virginia Press, 1916.

Warren, George F. and Pearson, Frank A., *Prices*, New York: Wiley & Sons, 1933.

Woodward, C. Vann, *The Strange Career of Jim Crow*, New York: Oxford University Press, 1955.

Auditor of the State of Arkansas, *Biennial Report*, Little Rock: State Printer, 1895–1911.

Auditor of Public Accounts of the State of Kentucky, *Report*, Frankfort: State Printer, 1866–85.

Auditor of Public Accounts of the State of Louisiana, *Biennial Report*, New Orleans: State Printer, 1891–1915.

Auditor of the State of North Carolina, *Annual Report*, Raleigh: State Printer, 1890–99.

Auditor of Public Accounts to the Governor and General Assembly of Virginia, *Annual Report*, Richmond: State Printer, 1890–1915.

Comptroller-General of the State of Georgia, *Report*, Atlanta: State Printer, various years.

Joint Committee on Tax Revision, Virginia, *Report*, Richmond: State Printer, 1914.

North Carolina Corporation Commission as a Board of State Tax Commissioners, *Report*, Raleigh: State Printer, 1900–15.

U.S. Bureau of the Census, *Negro Population, 1790–1915*, Washington: USGPO, 1918.

U.S. Department of Agriculture, Bureau of Agricultural Economics, "Prices of Farm Products Received by Producers: 3, South Atlantic and South Central States," *Statistical Bulletin No. 16*, Washington, 1927.

______, "Fluctuations in Crops and Weather, 1866–1948," *Statistical Bulletin No. 101*, Washington, 1951.

Accumulation of Property by Southern Blacks before World War I: Reply

By ROBERT HIGGS*

Examining data for five additional states, Robert Margo confirms my most important findings for Georgia: blacks accumulated property at a faster rate than whites during the late nineteenth and early twentieth centuries; and in the Deep South the dynamic profile of black accumulation reflected the fluctuating fortunes of the cotton economy. Analyzing cross-sectional county data for 1910, Margo obtains regression results for the additional states that differ somewhat from my findings for Georgia. He also raises the issue of potential bias in racial comparisons when assessed property values are used as proxies for true property values.

I. The New Data and Findings

The additional data compiled by Margo are a valuable contribution, and I do not doubt their trend. Some of their particulars, however, are peculiar. For example, as Margo's Table 1 shows, between 1895 and 1900 Arkansas whites had a sizable *decline* while blacks had a proportional *increase* in real assessed wealth per capita. Why the divergence? The same question may be asked about Kentucky during 1870–80 and Virginia during 1900–10. One reasonably supposes that, in a particular state during a given period, the same or similar events (for example, crop prices and yields) affected the accumulation of both races. Extreme racial disparities call the data into question. Other anomalies magnify one's misgivings. For example, during 1900–10, a generally prosperous period in Virginia and throughout the American economy, the real value of farm real estate, as reported to the census, increased by about 54 percent (Thomas Pressly and William Scofield, 1965, p. 44; George Warren and Frank Pearson, 1933, p. 13). Yet Margo's Table 1 shows that real white wealth per capita declined by about 8 percent during this decade in Virginia. I selected Georgia for my own study because, inter alia, its tax assessment system circa 1900 had been carefully evaluated by Laurence Schmeckebier (1900) and W. E. B. Du Bois (1901, 1906). I would feel more comfortable with the new data compiled by Margo if their sources had been subjected to similar scholarly scrutiny.

Margo is chary of the cross-sectional data for Arkansas and Louisiana (see his fn. 7). Flaws in the data may account for the poor regression results for these states. He notes that the F-statistics would have fallen short of the critical value for a 5 percent test if an additional "control" variable had not been added (see fn. 6). Whether one considers this respecification warranted or not, it makes these equations incomparable with the others presented in Margo's Table 2.

Only the equations for North Carolina are directly comparable (i.e., equivalently specified) with those for Georgia. In this comparison, only one variable (*PLANTATION*) has coefficients that are statistically significant and differ in sign between comparable equations. Otherwise, nothing in the North Carolina equations diametrically opposes the corresponding estimate in the Georgia equations. For three variables the coefficients are statistically zero in one equation and statistically nonzero in the comparable equation, but this is a weaker form of disagreement. Margo states that "few, if any, generalizations are possible" (p. 772). One should recognize, however, as he does, that the equations for Georgia and North Carolina agree about the statistically significant negative influences of expensive land and high illiteracy. On the former, the four Virginia equations, though not directly comparable, also agree.

*William E. Simon Professor of Political Economy, Lafayette College, Easton, PA 18042.

II. The Question of Assessment Bias

Margo raises the possibility that black property was assessed more heavily than white property. Regressing the county assessment ratio (for both races combined) on relative black population, total population density, and black illiteracy, he finds the coefficient of relative black population to be positive and significant. For the time and place under investigation, however, many socioeconomic variables are correlated with relative black population; so this finding may be spurious rather than a true indicator of assessment discrimination against blacks.

A more exacting test is possible if one be willing to consider data for a later date. G. H. Aull and Ernest Riley (1938) obtained information on 1,544 individual farms scattered across South Carolina's nine type-of-farming regions in 1935–36. The assessment ratios they found appear in Table 1. The table also shows a proxy, constructed by me from data in the 1935 federal *Census of Agriculture*, for the relative portion of black-owned farm acreage. The two variables have little association. Spearman's rank correlation coefficient for these data is negative (-0.38), but does not differ significantly from zero at the 5 percent test level. Because it uses better defined and measured variables, this is a stronger test than Margo's regression, though it is not directly comparable, as it pertains to a different southern state and a later date. It does enhance the suspicion that Margo's regression result is spurious.

Inspection of the Virginia data in Margo's Table 3 reinforces one's doubts. Why are the racial differences in assessment ratios greater at the lower property values? (Is it reasonable to suppose that discriminating assessors systematically exercised relatively more racial favoritism among holders of cheap properties than they exercised among holders of expensive properties?) Why would assessment discrimination be less in cities than in counties for cheap properties but greater for the highest-valued properties? Why is the racial assessment ratio unfavorable to *whites* for urban properties in the price range \$1,000–2,499? In sum, is the internal pattern of these data plausibly the outcome of racial

TABLE 1—ASSESSMENT RATIOS AND BLACK LANDHOLDING, SOUTH CAROLINA, 1935–36[a]

Region	Assessment Ratio on all Farms Studied	Portion of Farm Land Operated by Black Owners
1	22.9	2.43
2	25.9	6.01
3	20.5	6.79
4	29.2	6.13
5	28.0	9.12
6	26.5	7.21
7	27.4	6.36
8	20.3	12.73
9	16.0	12.67

Sources: Assessment ratio from Aull and Riley (p. 9); portion of farm land computed from data in U.S. Bureau of the Census (1936, pp. 482–83).

Note: The assessment ratio is measured as the ratio of assessed valuation to professionally appraised "normal market value." The latter was shown to be virtually identical to realized sales price for the state as a whole (see Aull and Riley, p. 9, fn.). The portion of farm land operated by black owners is measured as the entire acreage operated by black full and part owners as a percentage of all farm land.

[a]Shown in percent.

assessment discrimination? It does not appear to be.

That the assessment ratios vary inversely with the sales prices is an aspect commonly exhibited by real estate data. Levis Kochin and Richard Parks have shown that this relationship is a spurious indicator of vertical assessment inequity; it is a statistical artifact. A test in the form of a regression of sales values on assessment ratios "will indicate a bias, even when, by hypothesis, there is none" (1982, p. 516).

Aull, who directed extensive studies of farm real estate assessment in South Carolina during the interwar period, asserted that the relatively high assessment ratio on holdings of small acreage "is doubtless due in large measure to the relative importance of buildings on such properties but may be explained in part by the relative large number of such properties on the outskirts of cities and towns" (1940, p. 13). One can test the hypothesis about buildings with data from the 1925 *U.S. Census of Agriculture* (see Table 2). Clearly, the data are consistent with the hypothesis. In every southern census region,

TABLE 2—FARM SIZE AND RELATIVE VALUE OF BUILDINGS, SOUTHERN REGIONS, 1925

Farm Size (acres)	Ratio of Value of Buildings to Value of Land plus Buildings		
	South Atlantic	East South Central	West South Central
Under 20	0.35	0.35	0.31
20–49	0.28	0.27	0.21
50–99	0.28	0.28	0.19
100–174	0.28	0.26	0.17
175–499	0.26	0.23	0.15
500–999	0.21	0.19	0.13
1000 and over	0.13	0.15	0.07

Source: Calculated from data in U.S. Bureau of the Census (1927, p. 69).

an inverse relation obtained between the acreage of the farms and their ratio of building value to land-plus-building value. Hence, the higher assessment ratios observed on smaller holdings, where black owners were concentrated, reflected something other than discrimination by assessors against poor or black farmers.

Tipton Ray Snavely (1916), whose study seems to have provoked Margo's speculations about racial assessment discrimination, offered a different explanation, one reminiscent of Gary Becker's (1957) model of racial discrimination expressed by market price differentials. Snavely asserted that for several reasons, including "[t]o some extent, mere race prejudice, but to a larger extent other factors" (pp. 77–79), the selling price of black-owned real estate fell below the selling price of white-owned land of equivalent income-generating potential. The assessors, he believed, based their assessments on that potential without regard for the owner's race. Hence, the black assessment ratio was higher not because the assessors put its numerator up but because the market put its denominator down. The racial disparity in assessments, said Snavely, "is one in appearance rather than in economic fact" (p. 95).

Margo's discussion of assessment/census discrepancies is misleading. He states that my test of the reliability of the assessment data "implicitly defines the assessment ratio" and that there are "Errors in measuring the assessment ratio in this manner" (p. 774) that may account for the inconsequential results obtained when this "implicit" assessment ratio is added as an explanatory variable in regressions like those in his Table 2. But I never proposed nor intended that my tests of the reliability of the assessment data be interpreted to implicitly define an assessment ratio; therefore, the results of Margo's specification experiments hardly surprise me. But I would rather not take responsibility for something entirely of Margo's design.

Margo's perplexity about apparent discrepancies between assessment acreage data (which show *rural* land *owned*) and census acreage data (which show *farm* land *operated* by owners) is unnecessary. Much black-owned property was rented. About one-fifth of the rural acreage and about two-fifths of the rural land ownership units in Georgia circa 1900 were either rented or not used for farming (Du Bois, 1906, p. 526; Enoch Banks, 1905, p. 122; Thomas Woofter, 1920, pp. 57–58).

Some of the rural land assessed was forest land. In South Carolina, one authority ascribed a large part of the apparent vertical inequity in assessments to the timber holdings: "much of the timber in South Carolina is...[on properties of 500 acres or more], and the low average rate of assessment [on such properties] is indicative of the fact that, while no provision is made in the statutes for a classification of property for tax purposes, timber appears to enjoy a favorable position in this respect" (Aull, p. 13). The same kind of inequity may have prevailed in Georgia and other southern states.

Margo thinks that overassessment of blacks may have increased with time and hence that a substantial part of the apparent increase in black wealth could be spurious. I doubt this. No evidence is offered except the well known but questionably pertinent fact that certain forms of governmental discrimination against blacks (for example, segregation ordinances, voting restrictions, and school funding inequities) increased during the two decades after 1890. The numerical example, which Margo represents as illustrative, exceeds a plausible upper bound estimate of the effect of growing assessment discrimination. As he

recognizes, however, even this extreme calculation does not refute the main conclusion I reached in my 1982 article: black accumulation outpaced white accumulation and hence blacks closed the (relative) racial wealth gap inherited from slavery.

I disagree strongly with Margo's conjecture that "the bias in the Georgia case could be much larger" (p. 775), and I consider his illustrative calculations in footnote 15 highly implausible. The main reason for my disagreement is that Georgia, unlike Virginia, relied on a system of *self-assessment* by holders of rural properties (Schmeckebier, pp. 226–29; Snavely, p. 68). Are we to believe that Georgia's black property owners, in reporting the value of their rural property to the tax authorities, increasingly discriminated *against themselves*?

Another reason for doubting the significance of growing assessment discrimination, even in Virginia or other places where officials made the assessments, is that the blacks did not complain of it. They did complain vociferously about many other forms of discrimination: by police, courts, legislatures, prison authorities, school boards, labor unions, and various others. But I have never encountered a single allegation of direct racial discrimination by tax assessors. (Snavely, as noted above, denied direct discrimination by assessors. He claimed there was indirect discrimination via the market pricing process as well as via general, as opposed to race-specific, vertical and geographical inequities.) If such discrimination had been growing significantly more burdensome, we probably would have heard about it.

III. So What?

Notwithstanding some unresolved questions about the reliability of the data and the most warranted interpretation of certain causal relationships, Margo and I have brought forward a substantial body of evidence to sustain an important conclusion: during their first half-century of freedom, black southerners rapidly accumulated property and steadily increased their share of the region's real wealth. They did so despite many difficulties, including a virtually complete loss of political rights around the turn of the century. In my 1977 book and elsewhere, I have presented this experience as consistent with a hypothesis recently restated in a more general form by Thomas Sowell: "Historically, the relationship between political success and economic success [for an ethnic group] has been more nearly inverse than direct" (1983, p. 169). The facts adduced by me and by Margo on black property accumulation, taken together, lend credibility to this hypothesis.

REFERENCES

Aull, G. H., "Rural Land Holdings in South Carolina," *Bulletin 331*, South Carolina Agricultural Experiment Station of Clemson Agricultural College, October 1940.

______ **and Riley, Ernest**, "Some Inequalities in the Assessment of Farm Real Estate in South Carolina," *Bulletin 313*, South Carolina Agricultural Experiment Station of Clemson Agricultural College, January 1938.

Banks, Enoch Marvin, *The Economics of Land Tenure in Georgia*, New York: Columbia University Press, 1905.

Becker, Gary S., *The Economics of Discrimination*, Chicago: University of Chicago Press, 1957.

Higgs, Robert, *Competition and Coercion: Blacks in the American Economy, 1865–1914*, New York: Cambridge University Press, 1977.

______, "Accumulation of Property by Southern Blacks before World War I," *American Economic Review*, September 1982, *72*, 725–37.

Du Bois, W. E. B., "The Negro Landholder of Georgia," *Bulletin No. 35*, U.S. Department of Labor, July 1901, 647–777.

______, "The Negro Farmer," in U.S. Bureau of the Census, *Twelfth Census of the United States: 1900, Special Reports, Supplementary Analysis and Derivative Tables*, Washington: USGPO, 1906.

Kochin, Levis A. and Parks, Richard W., "Vertical Equity in Real Estate Assessment: A Fair Appraisal," *Economic Inquiry*, October 1982, *20*, 511–32.

Margo, Robert A., "Accumulation of Property

by Southern Blacks before World War I: Comment," *American Economic Review*, September 1984, *74*, 768–76.

Pressly, Thomas J. and Scofield, William H., *Farm Real Estate Values in the United States by Counties, 1850–1959*, Seattle: University of Washington Press, 1965.

Schmeckebier, Laurence Frederick, "Taxation in Georgia," in J. H. Hollander, ed., *Studies in State Taxation*, Baltimore: Johns Hopkins Press, 1900.

Snavely, Tipton Ray, *The Taxation of Negroes in Virginia*, Charlottesville: University of Virginia Press, 1916.

Sowell, Thomas, *The Economics and Politics of Race: An International Perspective*, New York: William Morrow, 1983.

Warren, George F. and Pearson, Frank A., *Prices*, New York: Wiley & Sons, 1933.

Woofter, Thomas Jackson, Jr., *Negro Migration: Changes in Rural Organization and Population of the Cotton Belt*, (1920), New York: Negro Universities Press, 1969.

U.S. Bureau of the Census, *United States Census of Agriculture: 1935: Reports for States with Statistics for Counties and a Summary for the United States*, Vol. I, Washington: USGPO, 1936.

______, *United States Census of Agriculture, 1925: Reports for States, with Statistics for Counties and a Summary for the United States*, Washington: USGPO, 1927.

Methodology: A Comment on Frazer and Boland, I

By ABRAHAM HIRSCH AND NEIL DE MARCHI*

Milton Friedman has been taken by some to be a follower of the logical positivists and by others a disciple of the philosopher Karl Popper, but until recently not enough has been said by way of support for such claims to make it possible to evaluate their substance. Recently, however, Laurence Boland (1979) has suggested that Friedman should rather be considered an "instrumentalist," meaning by this that for Friedman "theories are convenient and useful ways of (logically) generating what have turned out to be true (or successful) predictions or conclusions" (pp. 508–09) and Boland has given us a substantive analysis of his position. More recently Boland has combined forces with William Frazer (1983) to try to demonstrate that while Friedman is an instrumental thinker, and instrumentalism is a point of view against which Popper has argued vigorously, still Friedman's instrumentalism is in some sense compatible with Popper's falsificationist doctrine. These efforts are certainly in the right direction. A major reason there has been so little return on such a vast amount of discussion about Friedman's methodology is that the many participants in the debate have focussed on one or another aspect of Friedman's formulations without trying to determine what general point of view underlay his approach. Boland has dug deeper, and with Frazer has set out to show one of the fruits of this approach. The effort is therefore to be applauded for what it tries to do. But for reasons that will be given in this comment, Frazer and Boland wind up with a quite distorted picture.

First, a word about the genesis of the term "instrumentalism" as used by Frazer and Boland. They rely on Popper (1965, ch. 6) who has traced this way of thinking to the English philosopher George Berkeley. Popper includes within the term such philosophers of science as Ernst Mach, Heinrich Hertz, Richard von Mises, Moritz Schlick, and others. Their instrumentalism, in Popper's view, grew out of the problem of induction in the philosophy of science and was a way of rescuing the integrity of science while bypassing a troublesome problem, namely, that no one had found a way to prove deductively the truth of theories in the empirical sciences. The price paid for the rescue, as Popper sees it, is that empirical science was given a lower status than other realms of knowing (by making the sciences mere technologies), and this seems to be at least one of the reasons why Popper has rejected it. But Popper himself has not solved the induction problem so that he, in turn, pays a price for the rejection of instrumentalism, the price being the elimination of a substantial part of the empirical or observational element from science.[1] As he puts it, the alternative to induction is conjectural knowledge, whose objectivity lies in its being formulated so as to be open to rational criticism.[2]

There is, however, another type of "instrumentalism" which was developed by the pragmatic American philosopher, John Dewey (1938), who strongly influenced one of Friedman's mentors, Wesley Mitchell.[3] As

*Professor of Economics, Brooklyn College, City University of New York, Brooklyn, NY 11210, and Associate Professor of Economics, Duke University, Durham, NC 27706, respectively. The comments of Joop Klant on an earlier draft are greatly appreciated.

[1] So far as Popper is concerned, those who do the major work in science could be deaf, dumb and blind, so long as they had a few assistants with unimpaired senses to try to falsify the theories that scientists come up with. Facts are said to be the ultimate arbiter of the fate of a theory (Paul Schilpp, 1974, p. 971), but progress in science rests on the stage reached in critical discussion. Note that Popper (1972) claims he has resolved the induction problem, but Boland (1982) argues convincingly that he is mistaken in this.

[2] Popper (1972, pp. 9, 16, and passim).

[3] There are important instrumental components in both Mitchell's and Friedman's approach to economics though both deviate from instrumentalism in very different ways. These differences are considered in our (1983) paper.

has often been observed, Dewey's philosophy very well reflects the practical American environment, and having this orientation, it is not much concerned with the "rationalistic," nor with deductive proof. As a consequence, the very attempt to find a deductive proof of induction Dewey considered a will-o'-the-wisp arising out of a "quest for certainty" which is unattainable in this world (see Joseph Ratner, 1939, pp. 275–91). To Dewey the problem of induction is a pseudo-problem dreamed up by philosophers and best left alone. What is needed, he believed, is a theory of inquiry, based upon detailed and careful observation in the first instance, which shows how logical forms are generated in the process of inquiry, everyday as well as scientific, and how the scientist, as well as the layman, judges whether progress has been made when inquiry is undertaken.

Having approached science in this way, Dewey came up with notions which differ in some important respects from both Popper's and those of the instrumentalists whom Popper criticizes, as well as having features in common with both. Like Popper and those whom he calls instrumentalists, Dewey puts the test of implications derived from theory at the very center of "science."[4] But unlike Popper and those whom he calls instrumentalists, Dewey emphasized the observational basis of science (pp. 108–10); he did not recognize a "higher" sphere than the practical one of trying to achieve the ends of social existence (Ratner, pp. 945–49); and he rejected logical consistency as the one true God of the intellectual realm.[5] Against this background some of the limitations of Boland's analysis and of the Frazer-Boland analysis become apparent.

Looking at Boland's original paper (1979) we note, first, that he deals with just one of Friedman's essays (1953, first essay) and only one aspect of the thesis there developed. He does not try to explain, for example, why Friedman has involved himself with very extensive statistical investigations (Friedman, 1957; Friedman and Simon Kuznets, 1945; Friedman and Anna Schwartz, 1963; 1982) whereas most others have not; why Friedman is less concerned with an overarching logical consistency (one general explanation) (Friedman, 1974); why Friedman is less enthusiastic about the simultaneous-equations (structural) approach than many other economists (Friedman 1951); why Friedman is unconcerned not to have come up with an overall testable (or falsifiable) macro theory as other macroeconomists have tried to do (see John Wood, 1981); why Friedman's approach to policy differs from that of others. As we show elsewhere (1983), all of these diverse strands can be accounted for as part of a general rationale once the nature of the basic point of view is properly identified. Second, Boland sees Friedman's as a "practical" approach. But does practicalism necessarily go with instrumentalism? Surely not the instrumentalism of Berkeley, Mach, and the rest whom Popper points to as instrumentalists. Thus, while Boland in this instance has certainly put his finger on an important aspect of Friedman's thinking,[6] he has failed to connect it to his instrumentalism thesis. We need the Dewey-type of (pragmatic) instrumentalism to enable us to make the connection.[7] Third, to Boland, Friedman's instrumentalism is built on a disjunctive type of argument (1979, p. 506), an approach, as he notes, which is more ap-

[4] Dewey was even critical of economists for resting the validity of their theories on assumptions (Dewey used the term "premises") and argued that they should instead attempt to test the implications (Dewey's term was "consequences") of the theory. See Dewey, pp. 504–06.

[5] In a letter (see Arthur Burns, 1952, p. 96), Mitchell noted that Dewey opened his eyes to the fact that systems of thought, from his grand-aunt's theology to Marshall, might be tolerably self-consistent but that this was not an adequate test of "truth."

[6] Though he certainly carries it too far. He—alone and in combination with Frazer—makes it appear as if science is no concern whatsoever to Friedman. This is not true. It seems rather that Friedman passionately wanted to contribute to economic *science*, because he felt, as Dewey did, that to be useful, economics had to be scientific.

[7] Pragmatism and instrumentalism do not necessarily go together. Thus, Frank Knight gave assent to the pragmatic view of William James (1935, p. 97) but was scathingly critical of Dewey's instrumentalism (1947, essay three).

propriate for politicians trying to win elections (p. 506, fn. 8) than scientists searching for truth.[8] This is inconsistent with the implication in Frazer-Boland to the effect that Friedman was influenced by Popper, for it seems to us that had Friedman been influenced even in a general way by Popper, he would have amended his practice in order to try to become a "true" (Popperian) scientist. Of course, this would not necessarily have occurred were Friedman's major concern technology rather than science, but all of the evidence points in the other direction. Not only does Friedman again and again talk about "economic science,"[9] but in the very article with which Boland is concerned he uses Galileo as a model, not a well-known engineer.

It is worth pausing at this point to remind ourselves of what it is that Friedman and Popper are trying to do, each in his own work. Friedman wants useable predictions of the consequences of changes in the economic environment: specific answers to concrete problems, such as the effect of a tax on the demand for some commodity; the consequences for prices of a sudden jump in the rate of growth of the money stock. Facts—generalizations of related facts—and measurement are important here; but arguments based upon plausible generalizations only hinder the process of clarifying what policymakers have to argue and decide about. The ends of positive economic science are achieved when one of many possible explanations of events can be shown to hold up better under a wide variety of circumstances. One may then have some confidence in it as a basis for positive prescription (see Friedman and Schwartz, 1963, p. 676). Theories give more or less precise answers to the problem at hand; can explain more or fewer events; may have been checked over a more or less extensive base of cases; and, accordingly, are more or less useful.

By contrast, for Popper, "Theories are true or false and not *merely* [useful] instruments" (1972, p. 80). Deep down he is concerned with a striving after true knowledge (Schilpp, p. 1022; Popper, 1972, p. 697). Truth here has to do with propositions which have so far resisted refutation by facts and which thus give us some tentative ground for thinking we have moved in the direction of the target class—admittedly ideal and unattainable—of all true statements (Popper, 1972, pp. 57 ff.). Popper's work has been an extended attempt to formulate a logic for science as the pursuit of truth which avoids the pitfalls of Inductivism and Conventionalism (terms well explained by Boland, 1979). The process is one of identifying all the consequences of a conjecture, both those which can be tentatively accepted as true and those which could be false. By severe testing of its vulnerable parts—testing those strong claims for which matters could most easily be otherwise (Schilpp, p. 978)—and failing to reject them we pare down the falsity content and add to the tentative truth content. Thus truthlikeness is "a purely logical affair" and "*we test for truth*, by eliminating falsehood" (Popper, 1972, pp. 30, 81).

Friedman, too, rejects Inductivism and Conventionalism, as Boland (1979) makes clear.[10] Not surprisingly, Popper's theory of method issues in a methodological *rule* that is virtually indistinguishable from Friedman's. As if echoing Popper's point that science can only be "a competitive comparison of falsity contents" (1972, p. 81), Friedman, too, writes: "The greater the range of evidence that, if observed, would contradict a theory, the more precise are its predictions and the better a theory it is *provided it is not, in fact, contradicted*" (1974, p. 134). This agreement, however, is quite superficial. For Popper a rule for judging good scientific practice is a by-product of his philosophical

[8]In fact, as we read Boland, the politician is in a higher realm than Friedman because, according to Boland, the politician uses disjunctive reasoning where not all, but at least *one*, premise or assumption must be true for the argument to stand up logically; in Friedman's instrumental approach, however, according to Boland, *none* of the assumptions need be true.

[9]As Mitchell did—see, for example, Friedman (1953, pp. 282–83.)

[10]In an earlier article (1970), Boland argued that Friedman was a conventionalist but he has apparently changed his mind.

concern, namely, what would a theory of the growth of true knowledge have to look like to avoid Inductivism and Conventionalism. Friedman for his part is mainly a consumer of methodological insights. A good theory, for him, is first and last a *useable* theory;[11] if it happens also to conform to certain Popperian criteria of how good science is done that is an incidental bonus picked up along the way. One would be hard put to argue that Friedman's economic practice and its results have been driven by anything remotely resembling Popper's concern with a logical analysis of the growth of true knowledge. It is precisely by ignoring the different nature of the concerns of Popper the philosopher and Friedman the positive economic scientist that we end up with distortions such as those we alluded to earlier in Frazer-Boland. We deal here with five of these.

Notice, first, that Friedman has not only argued that "The theorist *starts with* some set of observed and related facts, as full and comprehensive as possible" (1953, p. 282, italics added), but in his labors for the National Bureau he has very conscientiously put this tenet to work. This, however, is the very antithesis of Popper's notion of how knowledge is acquired. Popper starts with hypotheses. Friedman's way could be called positivistic, since it has its roots in the idea that knowledge is acquired through the senses, but it is also an important element in Dewey's instrumental philosophy. Friedman, of course, was not ready merely to observe; he insists that observation be done in such a way that it has theoretical relevance. But he has tried to steer a middle course between empiricism and rationalism,[12] just as Dewey did.[13] Popper, however, on his own admission is not very far from the rationalists. He even calls himself a critical rationalist.

Second, as Frazer-Boland observe, and as Boland elsewhere argues at length (1979, Sec. 3), Friedman is not engaged in testing to falsify; he is busy trying to find explanations of successful outcomes (predictions that work; empirical regularities that he knows to hold). At the level of his basic concerns, in other words, he is busy with confirming success, not with refutation. This has to do with his starting with observation and with his interest in useable predictions. It is precisely what we would expect a Deweyan to be busy with. But it is not, as Frazer-Boland admit, the Popperian way. Popper enters only in so far as Friedman's criteria for choosing among possible explanations of success are, as we have seen, in agreement with his. The difference between testing for truth and comparing theories as instruments of prediction stands.

Third, to bring Friedman closer to Popper, Frazer-Boland argue that Popper's way of thinking, like Friedman's, is oriented around problems. But there is a difference in the way the term "problem" is used in the two cases. The kind of problems with which Friedman is concerned, as Frazer-Boland show, are practical social problems. This is Dewey's way. The major problems Popper is concerned with are cosmological, as shown, for example, in his views about the origins of modern science (1965, ch. 5). Starting problems for Popper may be practical problems (what can be done...?); but unless these lead to theoretical problems (*why* are...?) they are not "problems of explanation," which are the problems that really interest him (1972, p. 263). It seems to us, therefore, that there is also a very large difference here.

Fourth, Frazer-Boland present Friedman as a "short run" instrumentalist. What they seem to mean by this is that Friedman is not concerned at all with economic science, but rather with economic technology (i.e., de-

[11] This is not to say that any sort of theory will do for Friedman. He has a strong preference for theory that fits with the unifying themes of conventional price theory.

[12] Note, for example, Friedman's observation that rationalism on the one side and empiricism on the other

> ...have produced real and valuable improvements in the formal language available for describing economic interrelationships and in our detailed knowledge of the phenomena to be explained by economic theory. But they have left something of a vacuum in the equally vital intermediate area of theories or hypotheses that have implications about important phenomena susceptible of contradiction through observation. [1952, p. 457]

[13] This quest is at the very heart of Dewey's philosophy, particularly as it relates to social inquiry (see, for example, Dewey, ch. 24, particularly pp. 503–12).

vices that work in the short run) (see Boland, 1982, p. 152). This points back to our third reservation about Boland's 1979 paper, where we observed that Friedman has insisted that he is concerned with economic *science*, a claim which, from his Deweyan perspective, is true. Only it happens to be something quite distinct from Popper's science as the growth of true knowledge.

Finally, to make Friedman a Popperian it is not enough to force Friedman over to Popper; Popper also must be moved over to Friedman. To accomplish this task, Frazer-Boland give us the following quotation from Popper:

> ...For instrumental purposes of practical application of theory may continue to be used even after its refutation, within the limits of its applicability: an astronomer who believes that Newton's theory has turned out to be false will not hesitate to apply its formalism within the limits of its applicability....
> [1965, p. 113]

Using these notions for the purposes at hand raises a whole host of questions. For one thing, while technological methods involving falsified theory may work, scientific theory is more reliable. Even Friedman wants explanatory rationales for what works. Technological devices have use only because they save thought or computation. It surely cannot be said that Friedman prefers instruments, if he does so at all, because they save on computation. Nor does a study of his counterfactual analysis with Schwartz in the *Monetary History*, for example, suggest a lazy mind.

There are more subtle difficulties, too, in applying to Friedman the argument suggested by the above quotation from Popper. We say this because the way in which Frazer-Boland use the quotation makes it appear that they are implicitly postulating a tradeoff between immediate practical usefulness and long-run advance in economic science, and that they see Friedman choosing the former while most other economists opt for the latter. Our problem with this postulated tradeoff is, first, that a comparison of Friedman's work in positive economics with that of others does not bear it out. Second, there is little evidence that Friedman actually recommends such a choice. It seems to us rather that Friedman, like Mitchell and Dewey, opts for the practical—that is, predictive power—partly because he feels that the order of scientific advance is the identification of empirical regularities first, with explanations following. Friedman, as we have noted, recommends that the theorist start "with some set of observed and related facts, as full and comprehensive as possible" (1953, p. 282), and in most of his own work in positive economics he actually follows this order. Furthermore, Friedman seems to hold the implicit belief that philosophers' ideals of science might never be attained in economics and we may therefore have to settle for something else. He tells us, for example (1953, p. 38), that it would be desirable to have a more general theory of monopoly and competition than Marshall's which would do for economics what the theory of friction has done for physics; but so long as we do not have one we had better make the most of what we have. This "second best" choice, however, is not one in favor of short-run practical gains over long-run scientific progress. On the contrary, it amounts to an acceptance of the realities for what they are, a major tenet of a Deweyan way of thinking. Considering how often methodologists have chided economists for not living up to some ideal such as Popper's,[14] and how little effect this has had, the Deweyan position on this score, which Friedman seems implicitly to hold, can be said to have a good deal to recommend it. The gist of all this is that Frazer and Boland's attempt to make a short-run Popperian out of Friedman misses the mark by a wide margin.[15]

[14]Some methodologists (for example, Blaug and Hutchison in Spiro Latsis, 1976) keep scolding economists for not living up to Popper's ideal. (Blaug refers to Lakatos' ideal which in this respect is related to Popper's.) Is not the major reason for this failing that Popper's ideal is utopian?

[15]Friedman is reported by Frazer-Boland to have said that his methodological position is consistent with Popper's. We have suggested the limited, almost incidental sense in which this is true. We do not see how Friedman could subscribe to Frazer-Boland's analysis of the connection.

With one observation of Frazer-Boland we do agree. They tell us that "Friedman's early orientation in methods had no immediate background in established economics" (p. 129). It follows that the orientation must have come for the most part from outside established economics. We have tried to show that it could not have come from Popper, nor even those whom Popper calls instrumentalists.[16] We believe that it came from the way of thinking that prevails in the American environment and that is so well reflected in Dewey's pragmatic instrumentalism.[17] In this comment, we have had room enough only to suggest why we have doubts about Frazer and Boland's interpretation. In a more extended study we undertake the constructive task of trying to show how an understanding of some Deweyan aspects of Friedman's way of thinking helps shed light on his approach to the methodology of positive economics.

[16]Frazer-Boland say that Friedman was inspired by Popper (p. 135) and they have a list of distinctive traits which they consider characteristic of Friedman's approach (pp. 135–39). But they fail to make any connection between these traits and either Popper's way of thinking or that of the Popperian type of instrumentalists.

[17]There was also, of course, a Chicagoan input, which lies outside our concerns here. Surprisingly, however, aspects of a Deweyan type of instrumentalism were not uncommon even in the Chicago economics department when Friedman studied there. For some references, see our earlier paper, fn. 5.

REFERENCES

Boland, Lawrence A., "Conventionalism and Economic Theory," *Philosophy of Science*, June 1970, *37*, 239–48.

______, "A Critique of Friedman's Critics," *Journal of Economic Literature*, June 1979, *17*, 503–22.

______, *The Foundations of Economic Method*, London: Allen and Unwin, 1982.

______ **and Frazer, William J., Jr.**, "An Essay on the Foundations of Friedman's Methodology," *American Economic Review*, March 1983, *73*, 129–44.

Burns, Arthur F., *Wesley Clair Mitchell: The Economic Scientist*, New York: National Bureau of Economic Research, 1952.

Dewey, John, *Logic: The Theory of Inquiry*, New York: Henry Holt, 1938.

Frazer, William J., Jr. and Boland, Lawrence A., "An Essay on the Foundations of Friedman's Methodology," *American Economic Review*, March 1983, *73*, 129–44.

Friedman, Milton, "Comment" (on Christ: "A Test of an Econometric Model for the United States, 1921–1947"), in *Conference on Business Cycles*, Universities-National Bureau of Economic Research Conference Series, No. 2, University Microfilms, 1951, 107–14.

______, "Comment" (on Ruggles, "Methodological Developments"), in Bernard F. Haley, ed., *A Survey of Contemporary Economics*, Vol. 2. Homewood: Richard D. Irwin, 1952, 455–57.

______, *Essays in Positive Economics*, Chicago: University of Chicago Press, 1953.

______, *A Theory of the Consumption Function*, Princeton: Princeton University Press, 1957.

______, "Comments on the Critics," in Robert J. Gordon, ed., *Milton Friedman's Monetary Framework*, Chicago: University of Chicago Press, 1974, 132–85.

______ **and Kuznets, Simon**, *Income from Independent Professional Practice*, University Microfilms, National Bureau of Economic Research, 1945.

______ **and Schwartz, Anna J.**, *A Monetary History of the United States, 1867–1960*, Princeton: Princeton University Press, 1963.

______ **and** ______, *Monetary Trends in the United States and the United Kingdom: Their Relation to Income, Prices and Interest Rates, 1867–1975*. Chicago: University of Chicago Press, 1982.

Hirsch, Abraham and de Marchi, Neil, "American Pragmatic Instrumentalism and the Methodology of Positive Economics," mimeo., 1983.

Knight, Frank H., *The Ethics of Competition and Other Essays*, New York: Harper and Brothers, 1935.

______, *Freedom and Reform: Essays in Economics and Social Philosophy*, New York: Harper and Brothers, 1947.

Latsis, Spiro, *Method and Appraisal in Economics*, Cambridge: Cambridge University

Press, 1976.

Popper, Karl R., *The Logic of Scientific Discovery*, New York: Basic Books, 1965.

______, *Objective Knowledge: An Evolutionary Approach*. Oxford: Clarendon Press, 1972.

Ratner, Joseph, *Intelligence in the Modern World: John Dewey's Philosophy*, New York: Modern Library, 1939.

Schilpp, Paul Arthur, *The Philosophy of Karl Popper*, The Library of Living Philosophers, Vol. XIV, Books I and II, La Salle: Open Court Publishing, 1974.

Wood, John H., "The Economics of Professor Friedman," in George Urwick and James P. Quirk, eds., *Essays in Contemporary Fields of Economics: In Honor of Emanuel T. Weiler* (1914–1979), West Layfayette: Purdue University Press, 1981, pp. 191–241.

Methodology: A Comment on Frazer and Boland, II

By KEVIN D. HOOVER*

In a recent issue of this *Review* (1983), William Frazer and Lawrence Boland present Milton Friedman's methodology as instrumentalism. My purpose is not to question Frazer and Boland's interpretation of Friedman; rather it is to question their accompanying assertion that instrumentalism is a sound methodology for short-run, practical policy purposes.

Boland has set out his views on instrumentalism in greater detail elsewhere (1979; 1980; 1981; 1982, ch. 9). Since Frazer and Boland's "Essay" draws heavily on the earlier works, I will refer to them as well in order to fill in the implicit details of the argument. The larger purpose of the "Essay" is to examine the relationship between the methodologies of Friedman and of Karl Popper. Similarly, the larger purpose of Boland's earlier article (1979) is to defend Friedman's views from what Boland regards as misguided criticism and to provide a properly drawn "fundamental critique" of Friedman's instrumentalism. While it is important to recognize Frazer and Boland's larger purposes, I am not directly concerned with them, but rather with their analysis of instrumentalism itself. Most of what they say about instrumentalism refers to Friedman. I must assume, however, that what they say about Friedman *as an instrumentalist* applies to their understanding of instrumentalism generally.

According to Frazer and Boland, Friedman's essay (1953) calls "attention to the great relevance of positive economics (empirical study) for normative economics ('what ought to be')...." They continue, "...the question was which policy should be selected. Such question of how we get from where we are to the policy goal was thus seen as an empirical question, as one of selecting the most useful theory among available competitors" (p. 130). The promise of instrumentalism to Frazer and Boland is that it provides an effective method for answering this question. It does so by dissolving or ignoring the problem of induction and is as a method "...free from logical errors" (p. 131). Boland puts the case even more strongly: "Friedman's methodological position is both logically sound and based on a coherent philosophy of science—Instrumentalism" (1979, p. 503). (Compare Boland 1979, p. 521; 1980, p. 1557; 1982, p. 151.)

Boland qualifies his support of instrumentalism in an important respect: "Instrumentalism is always limited to short-run practical problems. If one is looking for a more universal understanding of the workings of the economy—that is, a true theory of economics—then instrumentalism will never do, since it ignores the truth of theories" (1982, pp. 151–52). (Compare 1979, p. 521; Frazer and Boland, pp. 131, 141–42.) My contention is that, even for short-run, practical policy problems, instrumentalism is not a sound, effective methodology.

I. The Logic of Instrumentalism

The problem of induction is how can we validly infer a general proposition from any finite number of particular instances of it. There is no widely accepted solution to it.

Boland (1979, 1982) sets forth his position with the help of a classical distinction between two sorts of inference: *modus ponens* (A implies B; A is true; therefore, B is true) and *modus tollens* (A implies B; B is false; therefore, A is false). Each has an associated fallacy: *affirming the consequent* (B is true; therefore, A is true); and *denying the antecedent* (A is false; therefore, B is false). *Modus ponens* argues from *sufficiency* (A is a sufficient condition for B); *modus tollens* from *necessity* (B is a necessary condition

*Nuffield College, Oxford OX1 1NF, England. I thank David F. Hendry and Peter M. Oppenheimer for useful comments on an earlier draft.

for A). For *modus ponens* to be useful we must have criteria for the truth or acceptability of A; for *modus tollens* for the falsity or unacceptability of B. If A is a theory which includes some general statements, we would require induction to establish its truth; if B is a set of particular conclusions, Boland seems to think determining its truth to be less problematic.

In Boland's view, the instrumentalist proceeds in three steps (1979, pp. 509–12; 1982, pp. 143–48). First, because the instrumentalist believes that there is no inductive logic, he never uses *modus ponens*. Second, he reasons by *modus tollens* to rule out any theories whose consequences are contradicted by any particular facts accepted as true. Having narrowed the universe of theories down to those whose known consequences are true, he has no further use of *modus tollens*—it is merely a prefilter. Finally, he uses conventionalist criteria (for Friedman simplicity and fruitfulness) to select a theory to use for a particular occasion. This theory is free from logical error. It is *alogical*; the first two steps ensure that logic is barely relevant to it. We do not believe that we can establish true antecedents; therefore, we cannot use *modus ponens*. We restrict our attention to consequences known already to be true; therefore, we have no use for *modus tollens*.

But if we consider only consequences already known to be true, what do theories *do*? Boland's answer is that for the instrumentalist they are convenient generators of those consequences. They use "something like *modus ponens*" to do their work (Boland 1979, p. 511). Presumably, it is not *modus ponens* only because the instrumentalist considers the truth of the antecedents to be irrelevant (Boland 1979, p. 509; 1982, p. 142). It cannot be emphasized strongly enough that Frazer and Boland's claim to logicality for instrumentalism is successful *only if* we restrict our attention to already successful predictions—to consequences known to be true. In this case, however, theories as generators are not needed—at best, they are a convenient shorthand, neat summaries of particular facts already known.

II. The Irrelevance of Instrumentalism

So far Frazer and Boland's exposition cannot be faulted. But a crucial flaw enters with the claim that instrumentalism is a sound or useful methodology, in the sense that a practical economist would be warranted in adopting it for short-run policy purposes. This claim rests on an equivocation about predictions. Boland writes: "For Friedman, an instrumentalist, hypotheses are chosen because they *are successful* in yielding true predictions" (1979, p. 511, emphasis added). The present tense in this quotation masks the fact that the instrumentalist can choose a hypothesis only because it *has been successful*, while the whole interest in predictions arises from the hope that they *will be successful*. This ambiguity has important consequences for Boland's argument. He writes:

> So long as a theory does its intended job, there is no apparent need to argue in its favor.... For some policy-oriented economists, the intended job is the generation of true or successful predictions. In this case a theory's predictive success is always a sufficient argument in its favor.
>
> [1979, p. 508; 1982, p. 144]

The last sentence may appear to commit the fallacy of affirming the consequent. This is not so, however, if predictive success refers to *past* predictions already known to be successful. In that case, it merely states that the theory has passed the *modus tollens* prefiltering process (step two).

Unfortunately a (perhaps infinite) number of theories will pass the prefiltering process. As policy-oriented economists we want to predict what *will* turn out to be true. We want not past but future success at prediction. For this we must push beyond the class of consequences already known to be true.

Boland recognizes that past success provides no guarantee of future success (1979, p. 513). Frazer and Boland nevertheless believe instrumentalism is sound for short-run, practical policy problems. And their view of the

scope of the shorter run is generous:

> ...[P]hilosophically, Friedman's views on methodology are instrumentalist in the shorter run where policymakers reside (including in the philosophical context less than a quarter in duration, cycles, trends, and long swings). The span of time is in contradistinction to an infinitely distant future where some ultimately true theory may reside, as found in the imagination of some philosophers. [1983, p. 141]

Normative ends, policy choices, in such a short run are still about the future. Predictions about the consequences of policies cannot belong to the class of the already known to be successful. The only way a theory can be *logically* helpful in generating predictions in this as yet unexamined class of consequences is if it is supposed to be true. And this leads us once again right up to the problem of induction: what grounds are there for supposing a theory to be true?

III. Some Replies Anticipated

Frazer and Boland might argue that instrumentalism has no need for induction; as Boland puts it, "...instrumentalism is its own defense and its *only* defense" (1979, p. 522). The instrumentalist makes no assumption about the truth of his theory and, hence, has no need for justification. He is therefore free, as a practical matter, to use his theory until it fails him and will no longer pass the prefilter. But in this case, anyone is free to use almost any arbitrary theory. Even past predictive failure is no hindrance, so long as the theory asserts that the future is different from the past (see Nelson Goodman, 1965).

Even as a short-run, practical matter, we seek advice about the unknown future and, therefore, need to distinguish one generator of predictions from another. When theories advise incompatible actions, we must choose between them. For example, if one theory said hold $M1$ growth to zero in order to reduce inflation and the other said let it grow at 20 percent to reach the same goal, we could not do both.

Frazer and Boland (pp. 130, 142) recognize that such practical choices have to be made. They argue, however, that instrumentalism is particularly at home with such problems. But how to discriminate between theories? Boland answers: "From the stand point of instrumentalism, the only prescription is to choose the theory which is most useful" (1982, p. 148). Unfortunately for the instrumentalist, when we restrict ourselves to conclusions already known to be true, no theory is especially useful; and when we seek advice about the future, a *true* theory is ideally suited to give useful advice.

Against this Boland claims that a theory does not have to be true to be successful (1980, p. 1556; 1982, p. 151). It is enough that the observed consequences be *as if* the theory were true (Boland 1979, p. 513). But this is not enough. Useful predictions are generated by (something like) *modus ponens*. *Modus ponens* in the class of consequences not already known to be true requires not only that the antecedents be supposed to be true, but that they be in fact true. When one has success in prediction, it is *accident* not *logic* which leads to the success, unless the antecedents are either true or related to the truth in such a way that their falsehood is irrelevant (for example, as an approximation).

The inevitable counterexamples are, first, Newton's theory (Frazer and Boland, p. 142) —it is known to be false, but is still practically useful—and, second, the television repairman who believes that faulty transistors are caused by the death of little green men but has still managed to fix the set (Boland 1980, p. 1556; 1982, p. 145). Such examples do not begin to meet the objections.

Newton's theory is easily understood as an approximation. By approximation I mean that all the falsifying evidence falls outside a well-defined set of boundaries or levels of accuracy (for example, Newton's theory has only been falsified at velocities approaching the speed of light).[1]

[1] Boland's (1981) remarks on Herbert Simon's appeal to approximation do not apply here. Simon argues that

The green man theory might be isomorphic to a true theory—names do not matter. Even if it is not, the example has no force, since Boland considers only the case in which the television set in fact works after the repairman has looked at it, rather than the case in which we must choose who shall repair it. If the set works after it has been repaired, we have no need for any inference or method. We only need inference when the set is yet to be fixed, and here instrumentalism is no help at all.

Most logicians distinguish logical validity (a formal property of an argument) from soundness (a substantive property). An argument with false antecedents may be valid without being sound.[2] My position is that in the only cases in which instrumentalist arguments are sound, they are also jejune. The only sense in which restricting our attention to short-run, practical problems might make instrumentalism more logically warranted is if "*practical* problems" is construed to mean "*past* problems." For really practical problems this is useless.

Finally, a disclaimer. To point out that the problem of induction cannot be dismissed even for short-run, practical policy problems is not to offer a solution to it. I have not advocated conventionalism or any other methodology: I have merely pointed out that instrumentalism, while logically flawless in dealing with any problem to which we already have the answer, is not a logically sound or efficacious methodology for any problem that matters.

when assumptions are approximately correct, conclusions will be approximately correct. I define a theory to be an approximation when its limits of accuracy are either specified within the theory or known because it is "observationally nested" within a better theory (see W. H. Newton-Smith, 1981).

[2]The argument, All scientists are wise; Economists are scientists; therefore, All economists are wise, is logically valid. *If* the premises were true, the conclusion would follow. But it is unsound.

REFERENCES

Boland, Lawrence A., "A Critique of Friedman's Critics," *Journal of Economic Literature*, June 1979, *17*, 503–22.

______, "Friedman's Methodology vs. Conventional Empiricism: A Reply to Rotwein," *Journal of Economic Literature*, December 1980, *18*, 1555–57.

______, "Satisficing in Methodology: A Reply," *Journal of Economic Literature*, March 1981, *19*, 84–86.

______, *The Foundation of Economic Method*, London: George Allen and Unwin, 1982.

Frazer, William J., Jr. and Boland, Lawrence A., "An Essay on the Foundations of Friedman's Methodology," *American Economic Review*, March 1983, *73*, 129–44.

Friedman, Milton, "The Methodology of Positive Economics," *Essays in Positive Economics*, Chicago: University of Chicago Press, 1953.

Goodman, Nelson, *Fact, Fiction and Forecast*, 2nd edition, New York: Bobbs-Merrill, 1965.

Newton-Smith, W. H., *The Rationality of Science*, London: Routledge and Kegan-Paul, 1981.

Methodology: Reply

By WILLIAM J. FRAZER*

Much has been written in reaction to Milton Friedman's famous essay on methodology (1953). As Bruce Caldwell has said:

> [It] is probably the best known piece of methodological writing in economics. It is also a marketing masterpiece. Never before has one short article on methodology been able to generate so much controversy.... Yet ironically, the methodological prescriptions advanced in the essay have become widely accepted among many working economists. And this happened *without Friedman even having directly responded to his critics*!
>
> [1982, p. 173, emphasis added]

More recently, in the thirty years since the publication of the essay, the controversy has shifted in two stages. The first shift was made by Lawrence Boland (1979), and then the second by myself and Boland (1983).

The main initial shift in the controversy was to do several things: identify Friedman with an instrumentalist position, to view his famous essay as logically sound on its own terms, and to introduce the notions of *modus ponens* (passing truth forward) and *modus tollens* (passing error backward), as a part of the controversy over the truth of axioms versus the testing of theories against one another on the basis of predictions. The second shift was not a reversal in any sense of the first, but rather it extended discussion in two further directions: that is, (*i*) it identified Friedman with a variant of the mainline instrumentalist philosophy which Boland and I referred to as "piecemeal engineering" (p. 141), and (*ii*) it extended the meaning of this latter variant by examining what Friedman actually did in his research, in executing his short-run, policy-oriented methodology.

*Professor of Economics, University of Florida, Gainesville, FL 32611.

Under (*i*) and (*ii*) of the developments under the second shift, a greater flexibility was added to what may be a more formally structured philosophy of instrumentalism. Under (*i*), the role of theory as an explanation or argument in favor of a policy oriented statement is not omitted (Frazer-Boland, p. 142).

Under (*ii*), we have a philosophy that one can proceed with in dealing with day-to-day policy matters without having to wait for (or to claim) that the one true economic theory has been found. The policymaker and the economist in this context reside in the short run in the sense that action or inaction is expected within time frames of less than one quarter in duration, cycles, trends, and long swings. There will not be delay in the advancement of argument in favor of a theory until the one true theory is discovered in the very distant future in which the philosopher resides.

In extending the discussion of Friedman's essay, the participants apparently think that discussion of methodology is important, perhaps more so at some times than at others. Such a time may be when the prevailing economic theory has been found to be inadequate, and when we seek to go back to our roots for guidance. Tjalling Koopmans stated the matter thus: "As new work unfolds, as changes in emphasis occur in the objectives of economics, and as fresh tools come into use, the desire for exchange of views on methodology recurs irresistibly" (1957, p. 130).

Of course, Friedman himself has been less patient with the methodological discussion, which is probably why he never answered directly his critics in the area in the first place. Expressing the view that the research speaks for itself in attempting to persuade others, Friedman concluded: "The process [the creative construction of hypotheses] must be discussed in psychological, not logical, categories; studied in autobiographies and

biographies, not treatises on scientific method; and prompted by maxim and example, not syllogism or theorem" (1953, p. 43).

We interpret this to mean that those following a successful line of inquiry will be studied and that their approaches will be emulated and followed. Along this line, it should be clear that the articles on Friedman's methodology and uses of methods by Boland and myself have not been pieces on methodology as such. They have instead been pieces on the clarification of Friedman's thinking as that is revealed to others by what he does in research, by his uses of methods, and by what he says about economic policy. That is why our earlier paper sought the extension of Friedman's early discussion of methodology as being important.

As to why Friedman may have never answered the critics of the early methodological piece we may conclude thus: the piece had set out a line of thinking that was to be followed in research that was planned for the next three decades; and the research could serve as an example of his thinking about the generation of fruitful empirical hypotheses. In recent correspondence, Friedman simply says, "I decided I could use my own efforts better in doing economics than in talking about how economics should be done" (March 7, 1984).

It is interesting in the light of the foregoing remarks to reflect that recent responses by critics have been directed at initial methodological discussion about instrumentalism per se (Abraham Hirsch and Neil de Marchi, and Kevin Hoover). This is in lieu of directing criticism toward the two directions in which the earlier article by Boland and myself extended discussion—notably toward the identification of Friedman's variant of instrumentalism, and the extension of the meaning of this variant by reference to what Friedman actually does in his research, writing, and public debates. In other words, with respect to that earlier article itself, there has been "an absence of flak."

REFERENCES

Boland, Lawrence A., "A Critique of Friedman's Critics," *Journal of Economic Literature*, June 1979, *17*, 503–22.

Caldwell, Bruce, *Beyond Positivism: Economic Methodology in the Twentieth Century*, London: George Allen & Unwin, 1982.

Frazer, William and Boland, Lawrence, "An Essay on the Foundations of Friedman's Methodology," *American Economic Review*, March 1983, *73*, 129–44.

Friedman, Milton, "The Methodology of Positive Economics," in his *Essays in Positive Economics*, Chicago: University of Chicago Press, 1953, 3–43.

Hirsch, Abraham and de Marchi, Neil, "Methodology: A Comment on Frazer and Boland, I," *American Economic Review*, September 1984, *74*, 782–88.

Hoover, Kevin, "Methodology: A Comment on Frazer and Boland, II," *American Economic Review*, September 1984, *74*, 789–92.

Koopmans, Tjalling, *Three Essays on the State of Economic Science*, New York: McGraw-Hill, 1957.

Methodology: Reply

By LAWRENCE A. BOLAND*

Many years ago in a philosophy journal I noted that "all economists can be divided into two groups—those who agree with Milton Friedman [1953, ch. 1] and those who do not" (1970, p. 239). Obviously I was quite naive in thinking that if Friedman's essay is important, it is worthy of competent criticism. Examining the major critiques I was quite disappointed. A careful reading of Friedman's essay from the perspective of Karl Popper's philosophy of science convinced me that none of the critics presented a logically adequate criticism. They merely criticized various invented caricatures of Friedman's essay. How could so many otherwise competent critics be so universally wrong?

This question posed an interesting puzzle. My response was to understand both Friedman and the critics on their own terms. The solution to the puzzle was to recognize that, in all cases, the critics' terms of reference were not the same as Friedman's. I presented my evidence in my 1979 article. First I argued, using Popper's terminology, that Friedman's essay was a logically consistent version of "instrumentalism." Second, on the basis of this view of Friedman's essay, I examined the adequacy of each of the major critiques. Such a strategy was risky in two ways. Not only would I have to correctly find where each critic failed to understand Friedman's essay, I must also represent accurately Friedman's essay. Failure to do so would invalidate my criticisms of the critiques. Fortunately, as Friedman has communicated to me and other methodologists, I was "entirely correct" in my characterization of his essay. So, it would seem that the only possible source of error could be my case against each of the critiques. Apparently, only one of the critics (Eugene Rotwein, 1980) could find anything with which to disagree. I replied by pointing out that this critic was again objecting to Friedman's essay solely because it did not provide a methodology acceptable to philosophers of a logical positivist persuasion. Since Friedman was not attempting to provide such a methodology, this type of critique will obviously be irrelevant (see my 1980 paper).

My 1979 paper closed with a statement of requirements for any effective critique of Friedman's 1953 essay. To convince Friedman's many followers, such a critique must be in terms of the objectives of that essay. It is pointless to criticize Friedman's essay for not achieving goals which he has never endorsed or which he has even denied. Since my 1979 essay appeared, it would seem that almost everyone accepts my argument that Friedman's essay is nothing more than an instrumentalist defense of instrumentalism. Nevertheless, none of the recent critics fully grasp the obvious line of instrumentalist-type criticism which would examine the domain of problems for which Friedman's instrumentalism *is* applicable.

It is not yet widely appreciated that Friedman's essay, while an expression of instrumentalism, is thereby a rejection of an influential and excessively philosophical viewpoint—the one which Bruce Caldwell calls "positivism" or which Popper call "conventionalism" (see Caldwell, 1982, and myself, 1979, 1980, and 1982). For that matter, it should be now recognized that Friedman rejects all views of methodology that are intended to provide an algorithm or formula for finding the one true theory of economics. Regardless of how many readers misinterpret Friedman's references to John Neville Keynes's discussion of "positive science" (see my 1979 paper, p. 510), Friedman's essay is not very philosophically minded—it is a rejection of any need to deal with

*Simon Fraser University, Burnaby, B.C. Canada V5A 1S6. I wish to thank J. Addison, J. Chant, S. Easton, D. Hammes, S. Kamath, C. Reed and Z. Spindler for their comments on an earlier draft.

most of the questions which philosophically minded economists have been so concerned.

Many papers have been written in response to my 1979 paper. Most of them claim that I did not understand Friedman's argument. Given Friedman's defense of my understanding, I think they have to come up with a much more fruitful line of attack. Unfortunately, most of the recent critics appear to think that if one does not criticize Friedman's essay, then one *must* be supporting Friedman, I reject this dichotomy. Nowhere in my 1979 essay is there any implication of support for Friedman's essay. The question now is, why do so many writers think that any defense of Friedman against unfair criticism constitutes support of Friedman's "methodology"?

I. The Sociology of Economics vs. Defeatist Conventionalism

The answer may be contained in my original 1970 observation quoted above. Most of the critiques come down to one sociological question: "which side are you on?," leaving no room for those of us who choose neither side if the choice remains only between Friedman's essay and conventionalist methodology. One does not have to pick sides in order to understand the widespread influence of or support for Friedman's essay. From the standpoint of methodology that Friedman has claimed that he is a follower of Popper's philosophy of science is an interesting puzzle, one pointed out by William Frazer. While Frazer and I may differ in our responses to the question of whose side we are on, I think we did adequately determine the extent to which Friedman's claim can be sustained.

Methodologists must meet the challenge of specifying the orthodox non-Friedman alternative. Samuelson's dictums stand out as an obvious alternative. Indeed, most economists probably hold Samuelson's views to be the only alternative. Unaligned methodologists find it interesting, that while there are numerous critiques of Friedman's 1953 methodological views, there is almost no criticism of Samuelson's methodological views anywhere in the leading journals (except two short articles—Donald Gordon, 1955; Stanley Wong, 1973). Since I think I have already shown (1979) that Friedman's methodology is logically consistent, acceptance of Samuelson's methodology cannot be based on any logical inconsistency of Friedman's essay.

It is perhaps even more puzzling to me that the methodological foundation of the widespread reliance on econometric model building is difficult to distinguish from methodological views presented in Friedman's essay. Econometric model building routinely accepts assumptions (equations) which are known not to be quite empirically true on the grounds that the true ones would be too complicated. As a consequence, proper econometric practice[1] can hardly be viewed as anything methodologically different from Friedman's "as if" approach.

II. Sound Methodology vs. Logically Sound Argument

Like many critics before, Kevin Hoover argues correctly that Friedman's essay is a poor substitute for a good, "sound methodology"—one that would be acceptable to the average logical positivist philosopher. As many before him, he misses the point. By rejecting the problems and concerns of those philosophers, Friedman's essay is not intended to be a good substitute. Hoover misses the point of my 1979 contribution in two ways.[2] Not only does he unfairly read my paper as "[my] support of instrumentalism," he erroneously reads the term "sound" to mean something more than "logically sound." Nowhere in my 1979 essay is the term sound used to mean "the premises were true" as Hoover seems to require. Nowhere in my paper with Frazer (1983) is it claimed that Friedman's essay is anything more than logically sound. Nowhere is any other meaning attached to the term sound other than logically sound. The only claim made for Friedman's essay was that it was without

[1] But not econometric theory which is usually based on conventionalism, see my 1982 study, chs. 7 and 8.

[2] Also, there are technical errors concerning logic and there are confusions of my views with my representation of Friedman's views. What is worse is that Hoover often presents my views as if they were his own!

logical errors. Hoover, it would seem, criticizes only his own caricatures of my 1979 paper and of Friedman's essay.

III. Popper vs. Conventionalist-Popper

While Abraham Hirsch and Neil de Marchi present some interesting ideas about Dewey's version of instrumentalism, they have little to say about the view of instrumentalism which, as I noted above, Friedman accepts as a correct representation of his view. Unfortunately, most of their paper concerns Popper's views of science and learning and about which they are simply far off the mark. They continually interpret Popper as if he were a paradigm conventionalist. In particular, they argue that "Popper's work has been an extended attempt to formulate a logic for science..." (p. 784). There could be no better evidence that they do not understand Popper.[3] Moreover, they did not try to understand Friedman—except to see him as only a variant of Dewey.

More distressing about the criticism of Hirsch and de Marchi (as well as Hoover) is its implicit view of methodology that there must be a correct algorithm or recipe for constructing short-run practical policy, or for choosing the "best theory" in the long run. Friedman rejects such concerns. So does Popper! It is probably the one point on which Popper and Friedman are in complete agreement. Methodology for Popper (see my 1982 study, chs. 10 and 11), and perhaps even Friedman, is a matter of understanding *why* economists do what they do—it certainly is not a set of formula prescriptions. Anyone studying methodology in order to obtain a methodological recipe is bound to be disappointed by Friedman's essay or any careful reading of Popper.

[3]There are two reasons. First Popper's major work (1959, which is the English version of a 1934 book) simply began with an alternative "logic of science"—that is, he definitely has not been attempting to come up with one since his first work! Second, when Popper began his work, the terms "logic of science" and "scientific methodology" usually referred to inductive logic and methodology. Popper has always criticized the widespread use of those terms in ordinary philosophy of science literature because they embody "inductivism." Modern "conventionalism" which is the viewpoint of the Hirsch and de Marchi is but a sophisticated variant of the inductivism which Popper has always explicitly rejected (see my 1982 study, ch. 1).

REFERENCES

Boland, Lawrence A., "Conventionalism and Economic Theory," *Philosophy of Science*, June 1970, *37*, 239–48.

______, "A Critique of Friedman's Critics," *Journal of Economic Literature*, June 1979, *17*, 503–22.

______, "Friedman's Methodology vs. Conventional Empiricism," *Journal of Economic Literature*, December 1980, *18*, 1555–57.

______, *Foundations of Economic Method*, Boston: Allen & Unwin, 1982.

Caldwell, Bruce, *Beyond Positivism*, Boston: Allen & Unwin, 1982.

Frazer, William and Boland, Lawrence, "An Essay on the Foundations of Friedman's Methodology," *American Economic Review*, March 1983, *73*, 129–44.

Friedman, Milton, "Methodology of Positive Economics," in his *Essays in Positive Economics*, Chicago: University of Chicago Press, 1953, 3–43.

Gordon, Donald F., "Professor Samuelson on Operationalism in Economic Theory," *Quarterly Journal of Economics*, May 1955, *69*, 305–14.

Hirsch, Abraham and de Marchi, Neil, "Methodology: A Comment on Frazer and Boland, I," *American Economic Review*, September 1984, *74*, 782–88.

Hoover, Kevin, "Methodology: A Comment on Frazer and Boland, II," September 1984, *74*, 789–92.

Popper, Karl, *Logic of Scientific Discovery*, New York: Science Editions, 1959.

Rotwein, Eugene, "Friedman's Critics: A Critic's Reply to Boland," *Journal of Economic Literature*, December 1980, *18*, 1553–55.

Wong, Stanley, "The F-Twist and the Methodology of Paul Samuelson," *American Economic Review*, June 1973, *63*, 312–25.

Test of the Lemons Model: Comment

By MICHAEL D. PRATT AND GEORGE E. HOFFER*

In a recent article in this *Review*, Eric Bond (1982) provided an empirical test of the "lemons" model. Using the 1977 *Truck Inventory and Use* (*TIU*) *Survey*, Bond determined that there was not a significant difference between the maintenance records of pickup trucks which were purchased new and those purchased as used vehicles. He suggests several explanations for his findings. These include the counteracting institutions discussed by George Akerlof (1970), as well as the ability of consumers to obtain enough information to eliminate the buyer-seller informational asymmetry. While Bond's conclusions are interesting, we propose to provide a finer test of the lemons model in order to determine whether these conclusions accurately reflect conditions in the market for used pickup trucks. Our model introduces two refinements to Bond's analysis.

We believe that Bond was unable to verify the lemons model because of the lack of expenditure data in the *TIU Survey*. As he notes, the quality of a used vehicle depends "not only on the probability of maintenance but on the costliness of repairs" (p. 837). However, the relative cost of repairs cannot be captured to any significant degree in the maintenance categories of the *TIU Survey*. As shown in Table 1, the four defined maintenance categories range in average expenditure from $100 to $1,200. Therefore, Bond's tests of differences in frequencies of repairs between vehicles acquired new and those acquired used are not as fine a test as necessary to verify the lemons model.[1] Not only does this test fail to capture the relative cost difference between repairs, it cannot account for vehicles that have more than one major repair during the year—perhaps the true lemon.

TABLE 1—RELATIVE REPORTED MAINTENANCE COSTS FOR PICKUP TRUCKS, 1977

Brakes	$ 100
Rear End—Differential	200
Transmission	400
Engine	1200

Source: Automotive Service Council of Virginia (1983).

Moreover, the Bond model does not capture the complete character of a lemons market. As noted by Akerlof, it is more likely that the owner of a lemon will sell his vehicle than the owner of a "creampuff." For if a lemons market is presumed to exist, the owner of the creampuff vehicle will not be rewarded for the superior condition of his vehicle. Therefore a comparison of trucks acquired new versus trucks acquired used would identify only those vehicles for which the lemon characteristic is permanent.[2] One should compare trucks purchased used during a specific time period with those held during that time period, whether they had been acquired as new or used vehicles, in order to capture both the "permanent lemon" as well as the "transitory lemon." The latter vehicle is the recently purchased used vehicle which is repaired by the new owner. In the model developed below, we test the lemons hypothesis using the *TIU Survey* data taking into account the two elements of a lemons market that should be included for a finer test of this hypothesis.

I. The Model

To reflect the essential time element and market process described by a lemons model, the *TIU Survey* data were segmented into two transaction groups.[3] In the first group

*Virginia Commonwealth University, Richmond, VA 23284. We thank Eric Bond, Robert Reilly, and an anonymous referee for their helpful comments.

[1]We thank the referee for this interpretation.

[2]We thank Bond for this clarification.

[3]Bond's analysis involved pickup trucks aged 5 years or less. We have included all pickup trucks in our analysis.

we included only pickup trucks that had been purchased used within one year of a state's survey date. These are the vehicles which would be more likely to include lemons. Our second group included pickup trucks which had been acquired either as new vehicles or used vehicles in some previous period and kept by that present owner. If there is a lemons market, this group should contain relatively more creampuffs. Excluded from both groups were vehicles that were reported to be leased, to be not in use, or to have been sold during the survey period.

Although Bond's test required that he control for vehicle age and total vehicle mileage, our test does not dictate this type of control. This is important, for to exclude some vehicles from the analysis because of their age and/or mileage characteristics precludes a full modeling of a lemons market.[4] At any given time, the market consists of a spectrum of trucks having various age and mileage characteristics. These characteristics provide information to a prospective buyer of a used vehicle but they do not effect the buyer-seller informational asymmetry.

To capture the relative cost of the reported maintenance categories we used the cost data presented in Table 1.[5] For example, as shown, in 1977 the average expenditure on transmission maintenance per year was four times that of the average expenditure on brake maintenance.

If the market for used vehicles is a lemons market, we would expect that the average yearly maintenance incurred for the four major repair categories reported in the *TIU Survey* to be significantly different between the two vehicle groups.[6] A priori, if the lemons hypothesis is true, we would expect that the average expenditure on vehicles purchased used within our observation period to be greater than the average maintenance expenditures on vehicles in the kept group. Under the null hypothesis that the market for used vehicles is not a lemons market, there should be no difference in the mean repair expenditures between our transaction groups. Thus a test of the equality of the mean repair expenditures between these groups is performed to determine the acceptance of the null hypothesis.

TABLE 2—TESTS OF THE EQUALITY OF MEAN MAINTENANCE EXPENDITURES[a]

	N	Mean	Standard Deviation	Standard Error
Recently Transacted	2,395	$3.18	5.42	0.11
Not Recently Transacted	10,799	2.51	4.88	0.05

[a]Differences in the means are significant at the .01 level.

The difference in the means reported in Table 2 is significant at the 1 percent level.[7] Therefore we reject the null hypothesis, and given the ranking of the mean repair expenditures, we conclude that the market for used pickup trucks is a lemons market. Therefore arguments for the presence of Akerlof-type counteracting institutions, as well as explanations for a smaller difference in buyer-seller information, although possi-

[4]Bond's tests require that he segment the market for used vehicles using age and mileage as a technical device. In the market, there is a segmentation of vehicles into sets that are considered by buyers and sellers to be nearly perfect substitutes. This need not correspond to a technical segmentation by age and lifetime mileage. Hence, although all trucks in Bond's sample are considered in his tests, this technical segmentation may not allow a proper test of the informational asymmetry between buyers and sellers within the same market segment.

[5]Because of obvious ambiguity, the category used to report "other maintenance" was excluded from the analysis. We would expect that this category would include maintenance types whose expenditures would be lower than the average brake repair. Inclusion of this maintenance category does not affect the results.

[6]There is not a significant difference between the average age or average lifetime mileage of these two transaction groups. Therefore, our test does not merely reflect differences in repair records that one might expect to exist due to significant differences in the average age or average lifetime mileage of vehicles in the two transaction groups.

[7]A sensitivity analysis shows that the results were invariant with respect to the actual dollar amounts reported in Table 1. The analysis is dependent on the relative ranking of the reported expenditure categories.

bly correct for some used markets, may not be significant factors in this market.[8]

[8]A direct test of the impact of counteracting institutions on the market for used vehicles is provided in our earlier paper (1983).

REFERENCES

Akerlof, George A., "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, August 1970, *84*, 488–500.

Bond, Eric W., "A Direct Test of the 'Lemons' Model: The Market for Used Pickup Trucks," *American Economic Review*, September 1982, *72*, 836–40.

Pratt, Michael D. and Hoffer, George E., "The Efficacy of Used Vehicle Disclosure Laws," Working Paper No. 83-01, Department of Economics, Virginia Commonwealth University, 1983.

U.S. Department of Commerce, Bureau of the Census, *Truck Inventory and Use Survey, 1977*, Washington: USGPO, 1981.

Automotive Service Council of Virginia, unpublished data, Richmond, Virginia, 1983.

Test of the Lemons Model: Reply

By ERIC W. BOND*

The purpose of my 1982 paper was to test for differences in quality between original owner trucks and those that had been acquired used that could not be accounted for by differences in age and previous usage. Michael Pratt and George Hoffer (P-H) make several modifications of my tests by constructing a new measure of repair costs that is a linear combination of the frequency of repair information of different maintenance types and by choosing to compare recently traded trucks with those that were not traded recently. In introducing these changes, they argue that we should throw away information on age and lifetime mileage that is available to the buyer. I will show below that ignoring differences in previous use can lead to biases toward finding quality differences in a market where there is perfect information.

Adverse selection of the type described in George Akerlof's (1970) "lemons" model arises when products that are equivalent from the point of view of buyers are known to be different by the seller. If 3-year-old trucks with average usage vary widely in quality and the variations cannot be observed by potential buyers, then those of lower quality will obtain the same price as those of higher quality and adverse selection will result. However, the fact that a 4-year-old truck becomes too costly to operate and is sold by its owner, while a 3-year-old truck owned by the same owner is not, is not necessarily related to the market for lemons because the buyer can observe the difference in age and anticipate costlier repairs. The price paid for the truck can be adjusted for the expected difference in repair costs and the market price schedule for used trucks will be a hedonic price schedule with compensating differences for observable quality differences. Casual empiricism suggests that the market uses such information, since trucks decline in value with age and high mileage trucks sell for a discount. Trade takes place because the buyer's valuation on the truck exceeds that of the seller.

When differences in age and lifetime mileage are controlled for, I find no difference in maintenance between recently traded pickup trucks and those not traded for trucks less than 10 years old. For trucks more than 10 years old, which were not included in my earlier study, I find evidence of more frequent repairs among trucks that have been traded recently. Thus, the significant quality differences found by Pratt and Hoffer arises from the inclusion of very old trucks in their sample, and from their failure to control for differences in characteristics that buyers can observe.

I. Observable Quality Differences

My 1983a paper presents a model in which users with high utilization rates and high maintenance costs will have a comparative advantage in the operation of relatively new trucks when operating costs increase with age. In such a model, trucks will be passed from the highest intensity of use owners to the lowest intensity of use owners over their life, since the optimal trading ages occur earliest for those with the highest intensity of use.[1] Table 1 shows a comparison of lifetime mileage for trucks that have been recently traded (according to the P-H definition) against trucks that have not been traded. For trucks less than 5 years old, there is a significant difference in lifetime mileage between those trucks that were traded and those that

*Associate Professor of Economics, Pennsylvania State University, University Park, PA 16802. I thank J. Rodgers, M. Roberts, and H. Geerts for helpful discussions.

[1]This problem is treated in detail in my working paper (1983b), which also estimates the impact of user characteristics on the age at sale.

TABLE 1—MEAN USAGE AND REPAIR COST BY TRANSACTION GROUP

Age (Years)		Number of Observations	Life Mileage	Cost
1–5	Traded	1,272	51.6[a]	2.16
	Not Traded	7,350	45.5	1.95
6–10	Traded	755	83.1	2.98
	Not Traded	4,154	82.0	2.80
>10	Traded	1,012	116.7	4.05[a]
	Not Traded	4,913	107.6	2.78

Notes: Not Traded = Not purchased within 12 months prior to sample, Traded = Purchased used within 12 months prior to sample, Cost = Average Repair Cost ($100) by P-H measure, Life Mileage shown in 1,000s.

[a]Mean significantly higher at 1 percent level than that for other transaction group of similar age.

were not traded, which is consistent with the model of trade based on comparative advantage. Tests that compare maintenance costs in this group without controlling for lifetime mileage differences would be biased toward finding higher maintenance costs for traded trucks. The mileage differences are not significant at the 10 percent level in the older groups, so that differences in characteristics are not likely to be a significant source of bias in these groups.[2]

The last column of Table 1 shows the mean repair cost according to the P-H measure for each group.[3] The difference between maintenance costs for trucks less than 5 years old is not significant at the 10 percent level, even though there is no adjustment for differences in lifetime mileage. The tests in my 1982 paper were limited to trucks less than 5 years old, so my findings of no difference in quality for this group is not sensitive to the alterations made by Pratt and Hoffer. If being a lemon is a permanent condition, then the first trucks to be scrapped should be the lemons and the tests should be done before significant scrapping has occurred. On the other hand, if major repairs must eventually be undertaken on all trucks, then the owner may be able to anticipate these events and sell the truck just before a major repair is needed. The remaining two rows of Table 1 show the mean repair costs for trucks 6 to 10 years old and greater than 10 years old. The difference in maintenance costs is not significant at the 10 percent level for trucks in the former group, but becomes significant at the 1 percent level for trucks more than 10 years old.

II. Recently Traded Trucks

The second change made by Pratt and Hoffer is to look for lemons among recently traded trucks rather than among trucks traded at least once. The issue involves where to place trucks that have been traded at least once, but not within the previous 12 months. If sellers of trucks fail to make repairs on trucks before they are traded, but these repairs are not an indication of permanently lower quality, then the P-H measure will be preferable. However, suppose that trading occurs when the owner obtains information (not available to buyers) that his truck is of permanently low quality. Then the stock of

[2]As I found in my 1983a paper for purchases of tractor trailers, the buyers of used pickup trucks in the older groups tend to be households or small firms. Both buyers and sellers in the oldest age group tend to be one-truck households that use pickups for personal transportation, so that comparative advantage does not explain trade in the oldest age group.

[3]Letting P_i be the estimated frequency of maintenance of type i, my 1982 paper used tests of the hypothesis $P_i^U - P_i^N = 0$, where the superscripts denote used and new trucks respectively. Letting α_i denote the maintenance cost of type i from the survey cited by P-H, their tests are of the hypothesis $\Sigma_i \alpha_i (P_i^U - P_i^N) = 0$. Clearly this cannot capture differences in the cost of a particular type of maintenance between different trucks (since the mean repair cost is assumed for all owners who report doing major maintenance), which is an important source of variation between trucks.

TABLE 2—EFFECT OF BEING TRADED ON PROBABILITY OF MAINTENANCE

	Repair Type			
Age	Engine	Transmission	Rear Axle	Brakes
1–5	−.03	−.19	.02	−.26[a]
	(.09)	(.11)	(.15)	(.09)
6–10	.12	−.09	−.18	.06
	(.10)	(.14)	(.10)	(.19)
>10	.46[a]	.53[a]	.18	.13
	(.08)	(.11)	(.13)	(.08)

[a]Denotes significant at 1 percent level; standard errors are shown in parentheses.

trucks that have been traded will be all those trucks for which unfavorable information has been received at least once in the past. In such a case, looking only among recently traded trucks for lemons would bias against finding quality differences unless unfavorable information is being received every year and the low quality trucks are traded every year. Thus, looking at recently traded trucks is an additional, although not necessarily preferable, way of examining the hypothesis.

To examine the sensitivity of my earlier results to the alternative definition of a used truck, I have reestimated my original model using the P-H definition of used trucks. Since maintenance is a dichotomous variable, the relationship between maintenance and observable characteristics can be estimated using a logit model:

$$(1)\quad \log(P_i/1-P_i)=\beta_0+\beta_1 AGE + \beta_2 LM+\beta_3 TRADED, \qquad i=1,\dots,4;$$

where P_i = probability that maintenance of type i was performed, LM = log of lifetime mileage, and $TRADED$ is a dummy variable equal to 1 if the truck was purchased used within the last 12 months. The presence of a positive and significant value of β_3 indicates lower quality for used trucks that cannot be explained by observable quality differences. Equation (1) was estimated using maximum likelihood.

The results of these regressions are reported in Table 2. For trucks less than 5 years old, three of the estimated coefficients are negative (with one significant at the 1 percent level), indicating maintenance is less likely on recently traded trucks once the accumulated use differences are controlled. For trucks 6–10 years old, there are no significant coefficients and only two are positive. For trucks more than 10 years old, all are positive and two are significant at the 1 percent level. It should be noted that a similar result for trucks more than 10 years old is obtained if we compare original owner trucks with those that have been purchased used.

Use of the P-H maintenance cost variable in tests that control for differences in observable characteristics presents some econometric problems. The P-H variable is not dichotomous, but it is not a continuous variable either because it can take on only 16 values (resulting from the combinations of the 2 possible outcomes for each of the 4 maintenance categories). Thus, estimation of a maintenance cost relationship using ordinary least squares would suffer from problems of heteroskedasticity. The joint test of all maintenance cost variables is likely to lead to findings of lower quality among trucks where individual tests do not find lower quality in cases where the individual tests all yield positive but insignificant values of $\hat{\beta}_3$. An examination of the results of Table 2 for the 1–5 and 6–10-year-old groups, where the hypothesis of equal quality could not be rejected, do not indicate such a pattern. The joint test is therefore unlikely to yield results differing from those of the individual tests.

III. Conclusions

These results indicate that Pratt and Hoffer find used trucks to be of lower quality not because they have a "finer" test, but because they fail to adjust for observable quality

differences and include trucks that are more than 10 years old. The failure to adjust for differences in age and previous use was shown to lead to biases toward accepting the hypothesis that trucks that have been recently traded are of lower quality than those not recently traded, with the bias strongest in trucks less than 5 years old.

The finding that trucks more than 10 years old that have been traded have significantly higher frequencies of maintenance than similar trucks not traded seems to support the hypothesis that counteracting institutions play an important role in assuring product quality. It is probably the case that most trucks sold by dealers fall into the category of trucks (1–10 years old) where there are no quality differences between those traded and those not traded. Trucks in the over 10-year-old age group account for approximately one-third of all pickup truck transactions, but account for less than 15 percent of annual mileage. Furthermore, the fact that a significant portion of trucks (presumably those of lowest quality) have been scrapped by that age indicates that this finding is not consistent with the interpretation of a lemon as a truck of permanently lower quality.

REFERENCES

Akerlof, George A., "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, August 1970, *84*, 488–500.

Bond, Eric W., "A Direct Test of the 'Lemons' Model: The Market for Used Pickup Trucks," *American Economic Review*, September 1982, *72*, 836–40.

______, (1983a) "Trade in Used Equipment with Heterogeneous Firms," *Journal of Political Economy*, August 1983, *91*, 688–705.

______, (1983b) "The Supply of Used Assets and Frequent Trading," Department of Economics Working Paper, Pennsylvania State University, 1983.

Pratt, Michael D. and Hoffer, George E., "Test of the Lemons Model: Comment," *American Economic Review*, September 1984, *74*, 798–800.

The Homogenization of Heterogeneous Inputs: Comment

By J. PATRICK GUNNING*

In a recent article in this *Review* (1981), James Buchanan and Robert Tollison used a truncated neoclassical model to deduce the allocational effects of forced equal-pay schemes, including equal pay for (a) workers of the same trade, (b) workers in the same firm or industry, (c) workers who perform the same work, and (d) the minimum wage. The model that Buchanan and Tollison used is one in which firms hire a single unit of each type of input, each may be hired in competitive markets where their prices differ, each has the same marginal physical product within the firm, and there is substitution within limits among inputs of different types. They also assume only human inputs, since they are interested in equal-pay schemes. The purpose of this comment is to show that, depending on their meaning of "substitution within limits," their model is either inconsistent or incorrectly specified.

The key to understanding the Buchanan-Tollison model appears to lie with identifying a production function that corresponds to "substitution within limits." The early masters of marginal productivity theory recognized that two separate assumptions could be made about production coefficients. They could be assumed fixed or variable. The assumption of fixed coefficients meant that there could be no substitution among inputs in a production process. The assumption of variable coefficients was used to illustrate cases in which a change in the relative prices of inputs would prompt entrepreneurs to alter their production methods in order to substitute inputs whose relative prices had fallen for those whose relative prices had risen.

If one assumed variable coefficients, the equilibrium conditions and the various marginal equalities could be easily traced. If one assumed fixed coefficients, the marginal conditions could not be expressed. As a substitute for marginal productivity, however, Marshall and others introduced the concept of net productivity. This was not satisfactory to mathematical purists. Moreover, it left the door open to the possibility of haggling over a surplus; since product exhaustion would not occur even at the margin in the typical case. The assumption of fixed coefficients has sometimes been expressed in terms of lumpiness or indivisibility.[1]

It seems clear that Buchanan and Tollison do not use the term substitution within limits to mean variable coefficients. Suppose that, by chance, a firm that faced variable coefficients hired only one unit of each input of different types and that these inputs had identical marginal value products. Suppose further that the firm, industry, and economy were in general equilibrium. Now let there be a general reduction in the wages of one type of input, perhaps as a consequence of a preference by owners of that input for less leisure. The relative price of the input would fall. In neoclassical theory, the firm would attempt to substitute more of this input for other inputs and would alter its production methods to do so. In Buchanan-Tollison's model, however, this substitution is impossible by assumption, since only one unit of each input is hired. One must conclude that substitution within limits does not mean variable coefficients.

Now consider whether they mean fixed production coefficients. Although this appears to be their meaning, it will be shown that it is inconsistent with their analytical framework. Consider their graphical representation of a firm, which provides the reference for their verbal discussion. The relevant parts are reproduced in Figure 1. Note that inputs of different types are arranged in order of ascending supply prices. An input of one

*Texas Wesleyan College and University of Delaware, Newark, DE 19711.

[1]See John Hicks (1932) for a greater elaboration of these points.

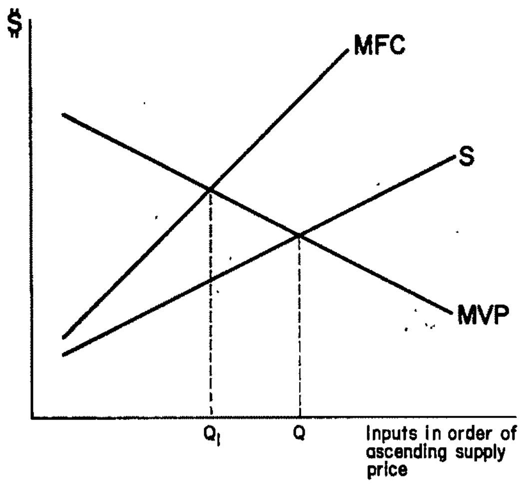


FIGURE 1

type which is available at a low wage would lie to the left of one of another type which is available at a higher wage. Note also the *MVP* curve. The implication of this curve is that a marginal input of a different type that is purchased at the next higher price will contribute less to the *MVP* than its predecessor. This implies a negatively sloped product demand curve. (In a different context it could imply diminishing returns; but in this context such an interpretation seems inappropriate. Buchanan and Tollison make no mention of diminishing returns.)

Assuming that input coefficients are fixed, the *MVP* function would necessarily be discontinuous, indicating that the firm purchased inputs in bundles. One could nevertheless use the concept of *MVP* loosely, so long as it was applied only to the bundle. Alternatively, the curve could be made continuous if the horizontal axis were labelled "bundles of inputs in ascending prices for hire."

This may seem like a minor point. Its importance arises from the fact that Buchanan and Tollison do not discuss bundles, and from their uses of the model. In their discussion, the firm is assumed to be able to discriminate in the wages it pays different inputs. This is a reasonable assumption since it is assumed to buy different inputs in different markets, where alternative employment opportunities are different. Roughly speaking, with fixed coefficients such a firm would hire bundles up to the point where the marginal value product of the bundle was equal to the marginal bundle supply price, which would be the sum of the wages of the individual inputs in a bundle. The equilibrium bundle could be indicated by Q on the graph.

Suppose that the firm faces an equal-pay law. Initially, I take this to mean that it must pay each input bundle the same total wage. Then, its decision to hire an additional bundle would depend not only on the cost of that bundle but on the increase in costs of earlier bundles as well. Roughly speaking, it would hire bundles only up to the point where the marginal value product of the bundle was equal to the marginal bundle factor cost. The equilibrium bundle on the graph could be indicated by Q_1. It is this idea that the graph seems to convey; although, as mentioned earlier, Buchanan and Tollison do not consider the bundling of inputs.

In their discussion, they want the equal-pay law to apply to all wage discrepancies. They consider a case in which all inputs are required by law to be paid the same wage. It is this intent that would render their analysis incorrect if they are assuming fixed input coefficients. If all the inputs within each bundle must be paid the same wage, the marginal supply price of bundles curve would shift upward in a parallel fashion. But Buchanan and Tollison do not mention this possibility.

Because there is so much discrepancy between the assumptions of fully variable or fully fixed production coefficients and Buchanan-Tollison's discussion, I ask whether they have some new type of production function in mind. They seem to believe that they do. They describe this as "team production" referring to Armen Alchian and Harold Demsetz's article (1972). The purpose of the Alchian-Demsetz model was to focus on the costs of monitoring employees and to show how, in a situation of fixed production coefficients, where the contribution of any single factor could not be determined by itself, monitoring was an important problem that

would give rise to the phenomenon of residual claimancy and thus one of the prime characteristics of the firm.

Buchanan and Tollison say nothing about monitoring. They refer instead to a construction team consisting of "a laborer, an excavator, a painter, a carpenter, a plasterer, a mason, a tile setter, an electrician, an insulator, a plumber, and a cabinet maker" (p. 33). In their discussion they point out that a house can be built with plaster or wood and that, therefore, plasterers and carpenters are substitutable. "But there is a point in this process of substitution beyond which the contractor-entrepreneur cannot go; he cannot build houses completely out of plaster" (p. 29).

Descriptions of this type are not very helpful and I can find no other indication that they even recognize the relationship between their assumption of team production and the two extreme neoclassical assumptions of fixed or variable coefficients. One might assume that they have in mind some combination of fixed and variable coefficients, in which case the analysis of this comment would have to be regarded as suggestive. I would have to conclude that, to the extent that inputs in Buchanan-Tollison's model are assumed substitutable, their model is inconsistent; and to the extent that inputs are not assumed substitutable, the model is incorrect.

REFERENCES

Alchian, Armen and Demsetz, Harold, "Production, Information Costs, and Economic Organization," *American Economic Review*, December 1972, *62*, 777–95.

Buchanan, James M. and Tollison, Robert D., "The Homogenization of Heterogeneous Inputs," *American Economic Review*, March 1981, *71*, 28–38.

Hicks, John R., "Marginal Productivity and the Principle of Variation," *Economica*, February 1932, *12*, 79–88.

The Homogenization of Heterogeneous Inputs: Reply

By JAMES M. BUCHANAN AND ROBERT D. TOLLISON*

What is the ultimate purpose of economic theory? If it is to help us understand reality, we must be prepared to apply the tools that have been developed for us to concrete problems. If we restrict our applications to cases where the tools fit perfectly, there would be precious little insight to be gained from economic theory. We should be ready to modify and stretch the tools we have, and invent new tools as required, if in so doing we can shed some light on observed practices in the economy.

This, as it were, is the well-known art of applying economics. For example, many institutions can be usefully analyzed with the theory of monopoly, though they do not strictly conform to the textbook definition of a monopoly. The same is true of any economic model. We thus care not one whit whether or not we may have, in our analysis of heterogeneous inputs, violated the virginity of the tools themselves.

The test of our earlier paper is simple: did it add to our understanding of the economy? J. Patrick Gunning has done nothing to suggest that the answer to this question is negative. We are happy to let him live in the escapist world of pure and unsullied models. There is nothing in his critique that offers any indication that these pure models can improve upon our explanation of the effects of treating heterogeneous inputs as if they were homogeneous.

Indeed, our discussion of the allocative effects of an equal pay for equal work regime hinges on little more than stylized common sense. We stressed that such a policy, promulgated in the face of differing opportunity costs for labor inputs, implies an allocative cost to the economy. What economic model does this simple application of the concept of opportunity cost violate? Moreover, many institutions are composed of an array of labor inputs with differing opportunity costs which are used to produce some output. Universities, as we noted in our original paper, immediately spring to mind, and it would seem that the effects of an equal pay for equal work policy in higher education are patently obvious and not disconfirming of our theory.

There seems to be little to be gained by a point-by-point reference to Gunning's misreading of our paper. One example will suffice. He states that we "...make no mention of diminishing returns," in spite of the fact that we do so in two places (p. 32) with specific reference to the point that he raises.

*Center for Study of Public Choice, George Mason University, Fairfax, VA 22030, and Center for Study of Public Choice and Department of Economics, Clemson University, Clemson, SC 29631, respectively.

REFERENCES

Buchanan, James M. and Tollison, Robert D., "The Homogenization of Heterogeneous Inputs," *American Economic Review*, March 1981, *71*, 28–38.

Gunning, J. Patrick, "The Homogenization of Heterogeneous Inputs: Comment," *American Economic Review*, September 1984, *74*, 805–07.

On Liquidity Preference—Again: Comment

By CASE M. SPRENKLE*

In their article in this *Review* (1983), Winston Chang, Daniel Hamburg, and Junichi Hirata (C-H-H) bemoan the neglect of the profession in recognizing that James Tobin's 1958 liquidity preference model does not fit the modern existence of a wealth of short-term, virtually (capital value) risk-free, assets which presumably dominate narrow money for speculative purposes. This criticism of the profession in general seems somewhat unfair. A cursory run through my not overly extensive library found substantial evidence that the profession is amply aware that narrow money is or may be dominated by short-term assets. As for books, Tobin himself in 1959, William Frazer and John Hicks both in 1967, Axel Leijonhufvud in 1968, Thomas Dernburg and Judith Dernburg in 1969, David Ott, Attiat Ott, and Jang Yoo in 1975, Charles Goodhart in 1975, and Jürg Niehans in 1978 all mention this domination. As for articles, Harry Johnson as early as 1961 and Martin Bronfenbrenner and Thomas Mayer in 1963 also raise the point. This list hardly suggests neglect of the problem, and in addition much of the verbiage in these works suggests the problem was widely known. However, C-H-H do have a valid point in that many popular current and recent macro and money textbooks develop Tobin's liquidity preference model uncritically without any discussion of its possible irrelevance to narrow money.

Chang et al.'s contribution lies in adding a third asset to Tobin's money and long-term bond model, the third asset being a short-term money market-type asset with low risk. They develop the model for a generalized utility function assuming normal probability distributions of returns and explicitly apply it using both quadratic and constant absolute risk aversion utility functions. They thus generalize the similar results of my 1974 paper which used only a quadratic utility function.

The results of these models are not only that narrow money is indeed dominated for speculative purposes as the risk on short-term approaches zero, but more importantly that narrow money is dominated even with some, not necessarily all that small, risk remaining on short terms. Since, as they show, actual risk on even the riskier short-term assets is very small indeed, this suggests that narrow money is dominated even for those economic units which either are prohibited from, or find it costly to make, investments in totally risk free short terms. Thus narrow money is well and truly dominated for speculative reasons. Or is it?

Whether or not narrow money is dominated depends on its characteristics and the existence and characteristics of short term money market assets. That is, it depends on institutional and financial details. A brief history is in order. When Keynes originally wrote, Treasury bills both in the United Kingdom and the United States existed, but only in small volume and the array of other short-term assets did not exist. For example, even by 1940 there were only \$1.3 billion of Treasury bills out of a total of \$34 billion of marketable U.S. government securities. Thus Keynes was reasonably accurate in viewing the financial world as consisting of an array of long-term assets, which could be conveniently represented by consols, and narrow money. Treasury bills in large volume, \$17 billion by 1945, are a creature of World War II, but low short-term interest rates after the war probably inhibited their use for speculative reasons. That is, narrow money was not so completely dominated. In addition the Fed pre-Accord policy of stabilizing long-term rates obviously reduced the extent of any speculative activity. After the Accord, short-term interest rates drifted up during the 1950's and with this drift came growth in

*Professor of Economics, University of Illinois, Champaign-Urbana, Champaign, IL 61820.

the types of alternative short term assets available. This upward drift widened the gap between short-term rates in general and the presumed zero rate on narrow money.[1] In addition, many of the new types of short-term assets were of the least risky or no-risk kind. Surely narrow money became dominated sometime during this period. Thus, in retrospect Tobin's liquidity preference article may already have been somewhat outdated by the time it was published.[2]

Ironically, at the very time that C-H-H appears, there is finally good reason not to bury Tobin's liquidity preference model, but to revive it in the proper context. Institutional and regulatory change have come full circle. The basis of my and C-H-H's notes is the assumption that narrow money yields no nominal return and thus is dominated for speculative reasons by the array of short-term assets which do, even if they also have some capital value risk. However, the introduction of interest-bearing NOW and ATS accounts and their inclusion as part of M_1 must lower the extent of dominance. And the introduction of Super NOW accounts in January paying money market rates should make part of narrow money a perfectly good substitute for long terms when capital values seem riskier.

It can be shown that in the same way that narrow money will be dominated as the variance (v_2) on short terms approaches zero if the return on money (r_1) is less than the return on short terms (r_2), short terms will be dominated as r_1 approaches r_2 and v_2 is small but positive. Using the explicit example of a quadratic utility function for intuitive "feel" and despite its well known problems, and using C-H-H's notation, optimal holdings of shorts and longs are

$$\text{(1a)} \quad A_2^* = \frac{(1+b)\left[(r_2-r_1)v_3^{.5}-(r_3-r_1)\rho v_2^{.5}\right]}{2bv_2v_3^{.5}(\rho^2-1)},$$

$$\text{(1b)} \quad A_3^* = \frac{(1+b)\left[(r_3-r_1)v_2^{.5}-(r_2-r_1)\rho v_3^{.5}\right]}{2bv_2^{.5}v_3(\rho^2-1)}.$$

From (1a), the condition for the domination of short terms by narrow money is

$$\text{(2)} \quad (r_2-r_1)/(r_3-r_1)<\rho(v_2/v_3)^{.5}.$$

Obviously with $\rho>0$, r_1 need not equal r_2 for short terms to be dominated, the extent of the necessary difference in rates depending positively on the relative risk on short terms.

To have narrow money held at all, but not necessarily dominating short terms, $A_2^*+A_3^*<1$ and from equations (1), the condition is

$$\text{(3)} \quad r_1>\frac{r_2(v_3-c)+r_3(v_2-c)}{v_2-2c+v_3} + \left(\frac{2b}{1+b}\right)\left[\frac{v_2v_3(1-\rho^2)}{v_2-2c+v_3}\right].$$

With v_2 (and thus c) equal to zero, r_1 must be greater than r_2, but with positive v_2 this is not so. As risk aversion increases ($b\to -1$), the condition becomes less strict. The effect of the covariance of returns is potentially mixed depending on the extent of risk aversion.

To have both narrow money and short terms in an optimal portfolio, the reverse inequality of equation (2) must hold as must equation (3).

These two constraints lead to a further constraint on r_2,

$$\text{(4)} \quad r_2>r_3+(2b/(1+b))v_3^{.5}.$$

The greater the variance on long terms and the greater the degree of risk aversion, the less strict is the condition for r_2.

[1]The exception to this is the concurrent growth of the compensating balance system of paying for bank services, lines of credit, and loans. The implicit interest rate paid by banks to larger, primarily corporate, customers did vary with market rates. However, most of these balances were held by nonfinancial corporations who are unlikely to have a Keynesian speculative demand for any financial asset.

[2]The first generation of modern postwar macroeconomics texts, such as Dernburg and D. M. McDougall (1960), used Tobins' then new liquidity preference model with some justification at least as a description of a fairly recent past. The publication of the first of the second generation texts (William Branson, 1972) with no recognition that the model was no longer realistic, prompted my comment in 1974. My comment obviously fell on deaf ears as shown by the contents of the wealth of new and revised texts since then, which presumably prompted C-H-H's comment.

It is clear then that either narrow money or short terms, or both, may be held in optimal portfolios if r_1 is close to r_2. In the case of Super NOW accounts, the interest rate has actually been above interest rates on some types of money market assets such as Treasury bills. Thus, banks holding a diversified portfolio of short terms are able and willing to offer higher rates on Super NOWs than on Treasury bills. And thus for many economic units too small to diversify efficiently their short-run portfolios, r_1 is actually above r_2 and narrow money dominates the relevant short-run asset.

There is one additional factor, which though neglected in the models, may be of substantial concern, and that is transactions costs. It is obvious that an investor wishing to switch out of long terms faces two transactions costs to switch into short terms, and only one transactions cost to switch into narrow money. Any effect of transactions costs obviously is to bias the decision toward narrow money. Although small, transactions costs will not be irrelevant if the expected duration of the switch is short as Hicks and others suggest.

It seems then, that liquidity preference is back, at least in a three asset form. Changing times in the form of high interest rates on some types of narrow money (along with, of course, concurrent changes in the definition of narrow money) have recreated a speculative demand for narrow money. Although the portion of narrow money now yielding money market rates is already of some significance, anyone familiar with the direction and pace of recent financial change would forecast further changes allowing for still larger portions of narrow money to bear higher interest rates. Given this, maybe we should not throw out our current textbooks, but instead make sure that those who correctly buried the theory in the past now know that it should be resurrected.

REFERENCES

Branson, William H., *Macroeconomic Theory and Policy*, New York: Harper and Row, 1972.

Bronfenbrenner, Martin and Mayer, Thomas, "Rejoinder to Professor Eisner," *Econometrica*, July 1963, *31*, 39–44.

Chang, Winston W., Hamberg, Daniel and Hirata, Junichi, "Liquidity Preference as Behavior Toward Risk in a Demand for Short-Term Securities—Not Money," *American Economic Review*, June 1983, *73*, 420–27.

Dernburg, Thomas F. and Dernburg, Judith D., *Macroeconomic Analysis, An Introduction to Comparative Statics and Dynamics*, Reading: Addison-Wesley, 1969.

______ **and McDougall, D.M.,** *Macroeconomics*, 1st ed., New York: McGraw-Hill, 1960.

Frazer, William J., Jr., *The Demand for Money*, Cleveland: World, 1967.

Goodhart, C. A. E., *Money, Information and Uncertainty*, London: Macmillan, 1975.

Hicks, John R., *Critical Essays in Monetary Theory*, London: Oxford University Press, 1967.

Johnson, Harry, "The General Theory after Twenty-Five Years," *American Economic Review Proceedings*, May 1961, *51*, 1–17.

Leijonhufvud, Axel, *On Keynesian Economics and the Economics of Keynes*, London: Oxford, 1968.

Niehans, Jürg, *The Theory of Money*, Baltimore: Johns Hopkins, 1978.

Ott, David J., Ott, Attiat F. and Yoo, Jang H., *Macroeconomic Theory*, New York: McGraw-Hill, 1975.

Sprenkle, Case M., "An Overdue Note on Some 'Ancient but Popular' Literature," *Journal of Finance*, December 1974, *29*, 1577–580.

Tobin, James, "Liquidity Preference as Behavior Towards Risk," *Review of Economic Studies*, February 1958, *25*, 65–86.

______, *Monetary Economics*, unpublished manuscript, 1959.

On Liquidity Preference—Again: Reply

By WINSTON W. CHANG, DANIEL HAMBERG, AND JUNICHI HIRATA*

Case Sprenkle's comment contains three main points:

1) Earlier criticisms of the James Tobin (1958) paper abound in the literature.

2) With a positive interest rate (r_1) on (narrow) money (A_1), there is a possibility that the portfolio demand for A_1 will be positive, provided the variance (v_2) of (algebraic) capital gains on short-term instruments (A_2) is sufficiently large.

3) In the case of Super NOW accounts, $r_1 > r_2$ (the interest rate on short term as exemplified by Treasury bills), and therefore A_1 may actually dominate A_2.

Point 1. Although we were unaware of some of the earlier criticisms of the Tobin model cited by Sprenkle, we did not claim none was made. On the contrary, we noted that "other authors have also expressed misgivings about the Tobin theory" (p. 420), and cited Robert Barro and Stanley Fischer (1976) as one example. Elsewhere, we remarked that our results are "obvious," but asserted the need of the obvious to be shown rigorously. We also noted "the approving presence of the Tobin model in so many macro and money texts in print" (p. 426).

Point 2. This point is implicit in the "portfolio" literature in general, and the literature on portfolio money demand in particular. The specific quadratic utility function employed by Sprenkle aside, this point simply says that given v_2, if r_1 is high enough, money (A_1) will be held. In other words, there exists a tradeoff between risk and return. Cast in this way, the point is not new. Tobin's two-asset model, for example, may be interpreted to say that with a large variance of capital gains from consols, money will be held even with $r_1 = 0$.

As it happens, Sprenkle's point is basically a complement to our footnote 6. The converse condition of this footnote can be put as follows: If $A_1 > 0$, then $\beta_2 \geq (-U_\mu/U_v)(r_2/2v)$. Thus, if A_1 is held, it is necessary that v_2 be sufficiently high. The above necessary condition can be satisfied even if $r_2 > r_1 = 0$. (In our model, $r_1 = 0$; in Sprenkle's case, $r_1 > 0$.) To verify the above condition, note that if $A_1 > 0$, then $\lambda = 0$ (by our condition (C)). Then our equation (6) implies that $U_\mu(\partial\mu/\partial A_2) + U_v(\partial v/\partial A_2) \leq 0$. Therefore, by equations (10) and (11), we have $U_\mu r_2 + 2U_v(A_2 v_2 + A_3 c) \leq 0$. Using the definition of β_2 (in footnote 6), we obtain $U_\mu r_2 + 2U_v \beta_2 v \leq 0$, which yields $\beta_2 \geq (-U_\mu/U_v)(r_2/2v)$.

Point 3. This point is an empirical one—a question of fact. We don't know the source of Sprenkle's information regarding interest rates on Super NOW accounts, but ours flatly contradicts his. Until October 1, 1983, the Thursday edition of *The New York Times* published a table in its financial section giving the rates on Super NOW accounts paid by 16 depository institutions in New York City and vicinity (New Jersey and Connecticut). On September 28, 1983, for example, a date when rates on Treasury bills ranged from 8.77 percent (for the one-week maturity) to 9.08 percent (for the 13-week bill), none of these institutions paid as much as the lowest of the bill rates. On that day, also, reported the *Times*, the "average rate paid... by 50 large institutions in the five largest U.S. markets" was 7.34 percent. A telephonic survey of the rates paid by eight institutions in the Buffalo area substantiated those reported in the *Times*.

This information should occasion no surprise. It is well known that depository institutions have been trying to keep the availability of Super NOW accounts a dark secret. Compared with the volume of advertising of the new money market deposit

*Chang and Hamberg: Professors of Economics, State University of New York, Buffalo, NY 14260; Hirata: Assistant Professor of Economics, Technological University of Nagaoka, Japan 949-54.

accounts (MMDAs), the advertising of Super NOW accounts has been minuscule. And with good reason: the latter permit costly unlimited checking, MMDAs only three checks, and a maximum of six transactions, per month.

Moreover, Sprenkle to the contrary notwithstanding, interest rates on Super NOW accounts, although market *related*, should normally be expected to be less than market rates. The reason is the reserve requirement (presently 12 percent) on Super NOW accounts (and all other checking accounts). *Ceteris paribus*, a reservable deposit should pay an interest rate equal to $i(1-r)$, where i is a market rate that banks pay for borrowed funds, and r is the legal reserve ratio on checking deposits. (Note that this spread is an increasing function of both i and r.)

Finally, note that Super NOW accounts share the same limited access (noted in our paper) as NOW accounts: both are denied to business firms, which have long been much the largest owners of checking accounts. In addition, Super NOW accounts have a severe drawback of their own: a required minimum balance of $2,500 for a checking account; this is calculated to be a potentially prohibitive requirement to most households.

All this is not to deny the use of Super NOW (and NOW) accounts as money. It is to question Sprenkle's effort to include them in the *portfolio* (speculative) demand for money. Because of their unlimited checking, they are better included in a *transactions* demand for money.

REFERENCES

Barro, Robert J. and Fischer, Stanley, "Recent Developments in Monetary Theory," *Journal of Monetary Economics*, April 1976, *2*, 133–67.

Chang, Winston W., Hamberg, Daniel and Hirata, Junichi, "Liquidity Preference as Behavior Toward Risk is a Demand for Short-Term Securities—Not Money," *American Economic Review*, June 1983, *73*, 420–27.

Sprenkle, Case, M., "On Liquidity Preference —Again: Comment," *American Economic Review*, September 1984, *74*, 809–11.

Tobin, James, "Liquidity Preference as Behavior Towards Risk," *Review of Economic Studies*, February 1958, *25*, 65–86.

Public Education: Comment

By GEORGE M. PERKINS*

In an article in this *Review* (1975), Martin Feldstein recommends a matching grant program for communities to achieve categorical equity in local education. Specifically, he supports a scheme which he purports will neutralize the influence of community wealth on local spending decisions. This grant mechanism varies the locally financed portion of expenditure directly with wealth. Such a program would be effective if demand for education is sufficiently price elastic. The price elasticity estimate obtained by Feldstein is elastic enough to suggest that manipulation of the tax price faced by a community through a matching grant scheme could achieve a desired level of categorical equity.

Several investigators, however, have obtained price elasticity estimates regarding the demand for local public education that question Feldstein's policy prescription. As shown in Table 1, these researchers have found this demand to be inelastic. In each of these cases, the matching grant necessary to achieve categorical equity may well have been too large to have passed favorably through the respective state legislature. Since most samples have been drawn from different places and at different times, it can be argued that the elasticity estimates reflect preferences that differ across space and time.

Helen Ladd and myself, however, like Feldstein, observe Massachusetts communities in 1970 and we both report price elasticity estimates that contradict Feldstein's conclusion that a matching grant scheme could be employed as a policy tool in achieving catagorical equity. In addition, we both incorporate state matching rates into their tax price variable. Moreover, despite the fact that each of us approaches the construction of this variable differently, Ladd allows for a partial shifting of the local tax burden away from the residential property owner and I introduced resources costs from service production, our price elasticity estimates are quite inelastic and different from Feldstein's estimate.

Motivated by this disparity among estimated price elasticities and by the implication their values have for the effectiveness of a wealth-neutralizing policy based on a matching grant program, this paper reports on 1) an attempt to replicate Feldstein's results, 2) an investigation of his expenditure relationship for other years (1973, 1975, and 1976), and 3) an analysis of expenditures by a panel of communities which is formed by combining time-series and cross-section data.

I. Cross-Section Estimates for 1970

As in Feldstein, the following linear expenditure equation is estimated:

$$(1)\qquad \ln E_i = \beta_0 + \beta_1 \ln W_i + \beta_2 \ln P_i + \sum_{j=3}^{K} \beta_j \ln X_{ji} + \varepsilon_i,$$

where E_i is education expenditure per pupil in community i, W_i is wealth per pupil, P_i is the net price that the community pays per dollar of expenditure on local education, and the X_{ji}s are other variables which reflect community resources and other characteristics.

Income and population data were published by the U.S. Bureau of the Census, local property values were provided by the Massachusetts Department of Revenue, while the remaining data were supplied by the Massachusetts Department of Education. Expenditures for education and equalized property values for each community were

*Assistant Vice President, First Wisconsin National Bank-Milwaukee, WI 53202, and Adjunct Associate Professor of Economics, University of Wisconsin, Milwaukee, WI 53201. I am indebted to the National Science Foundation and the University of Wisconsin-Milwaukee for financial support for my research.

TABLE 1.—COMPARISON OF ESTIMATED PRICE AND INCOME ELASTICITIES OF DEMAND FOR LOCAL PUBLIC EDUCATION[a]

Investigator	Sample	Estimated Elasticity: Income	Price
Barlow (1970)	52 Urban Michigan Districts (1960)	0.64 (0.10)	−0.34 (0.04)
Bergstrom et al. (1982)	2,000 Michigan Voters (1978)	0.64 (0.40)	−0.39 (0.25)
	469 Michigan Districts (1973)	0.38 (0.03)	−0.15 (0.02)
Black et al. (1979)	23 Delaware Districts (1973)	0.44 (0.09)	−0.15 (0.04)
Bradford-Oates (1974)	58 New Jersey Districts (1960)	0.65 (0.08)	−0.33 (0.07)
Brazer (1959)	40 Large U.S. Cities (1953)	0.73	−0.73
Feldstein (1975)	105 Massachusetts Cities and Towns (1970)	0.48 (0.07)	−1.00 (0.19)
Inman (1977)	58 Long Island Districts (1970)	0.61 (0.11)	−0.42 (0.28)
Ladd (1975)	78 Boston SMSA Districts (1970)	0.46 (0.15)	−0.48 (0.30)
Lovell (1978)	136 Connecticut Districts (1970)	0.65 [b]	−0.82 (0.14)
Perkins (1977)	38 Massachusetts Towns (1970)	1.02 (0.23)	−0.29 (0.17)

Note: Standard errors are shown in parentheses.

[a]A certain ambiguity exists among authors when reporting price elasticities in studies of the "Demand for Public Education." Here the price elasticities of education expenditure (per capita) are reported

$$d \ln E / d \ln P = d \ln PQ / d \ln P = 1 + d \ln P / d \ln Q.$$

[b]Lovell estimates the value of γ separately using it as a parameter in his demand equation, hence he does not report a standard error.

deflated by the number of public school pupils. The term $PRIV$ denotes the proportion of private school children in the population while PUP is public school pupils per capita. The matching rate, m, is calculated according to the formula mandated in chapter 70 of the Massachusetts General Laws. There are restrictions placed by law on this formula which effectively change the grant type from matching to unconditional. The former works by distorting price to induce a desired expenditure while the latter increments a community's resources. The donor expects that some of this increase will finance additional expenditure. If the municipal government receives a matching grant, then $P = (1 - m) < 1$ and no lump sum grant (SBG) is received. However, if any of the above mentioned limitations apply, $P = 1$ and a lump sum grant is received by the local government. Subventions from the federal government (FG) are also treated as lump sum grants.

Here the proportion of residential property in the local tax base (RES) and the pupil growth rate ($GROW$) are constructed with minor differences when compared with Feldstein's data.[1] Specifically, his residential proportion is calculated from a 1958 survey which he acknowledges may be inaccurate

[1]The data necessary to reconstruct these variables could not be obtained from the Departments of Education and Revenue. Responsible officials there speculated that over time the sought after information had been lost or discarded. Also, it appears the same has happened to Feldstein's data.

TABLE 2—A COMPARISON OF LOCAL DEMAND FOR EDUCATION, 1970

Results	Sample Size	*P*	*w*	*MEDINC*	*SBG*	*PRIV*	*PUP*	*FG*	*RES*	*GROW*	*INC*	$\bar{R}^2$
Feldstein	105	−1.000	0.283	0.475	0.066	−1.112	0.208	0.136	−0.118	−0.336		.64
		(5.41)	(7.45)	(6.99)	(5.08)	(2.82)	(2.84)	(5.04)	(2.74)	(4.42)		
Perkins	112	−0.704	0.444	0.947	0.013	−0.106	−0.266	−0.000	−0.202	−0.222		.81
		(3.45)	(9.08)	(11.47)	(2.62)	(3.88)	(3.48)	(0.05)	(2.69)	(2.16)		
	112	−0.669	0.46	0.880	0.011	−0.103	−0.282					.79
		(3.15)	(9.16)	(11.04)	(2.23)	(3.73)	(3.59)					
	341	−0.484	0.286		0.012	−0.044	−0.054				0.495	.53
		(2.70)	(11.33)		(2.44)	(3.12)	(1.05)				(8.79)	
	255	−0.471	0.262		0.009	−0.053	−0.116				0.706	.60
		(2.53)	(7.11)		(1.72)	(3.36)	(2.08)				(10.90)	
	143	−0.557	0.266		0.011	−0.081	−0.180				0.832	.74
		(2.72)	(5.16)		(2.10)	(4.20)	(3.37)				(11.89)	
	59	−0.722	0.524		0.013	−0.153	−0.087				0.477	.83
		(2.26)	(5.68)		(1.76)	(4.69)	(0.92)				(3.72)	

Note: Absolute values of t-statistics are shown in parentheses.

for 1970. Assessment data as of January 1, 1975 are used in this paper to construct *RES*. Feldstein's *GROW* is the ratio of students in 1970 to those in 1965 while here students in 1975 are compared to those in 1970.

Feldstein determined his sample size by eliminating small towns where data were not available for all variables.[2] Since the Departments of Education and Revenue report their series for all 351 Massachusetts communities, it is obvious that he truncated this population to a sample of towns with at least 10,000 inhabitants because census data on median family income (*MEDINC*) are not published for smaller communities.

Besides reporting results for 1970 alone, which are the focus for this paper, Feldstein also reported results based upon pooled data from 1965 and 1970. It appears that communities with less than 10,000 inhabitants in 1965 were also removed from his sample. A few other towns were removed from the sample because they participated in a regional school district. Finally, it was necessary to drop a few more communities in order to estimate Feldstein's log-linear equation because for these *PRIV* was not positive.

Following these steps yields a sample of 112 communities. Although there were 105 communities in Feldstein's sample, the price elasticity estimate does not change when the sample is reduced by considering only towns with larger populations. Furthermore, when median family income is replaced by per capita income (*INC*), which is available for all 351 cities and towns, the price elasticity estimates are robust throughout alternative sample sizes.

Table 2 compares the estimated elasticities obtained here with those reported by Feldstein. Two results are obvious: first, all price elasticity estimates reported here are lower than that obtained by Feldstein (−1.00); second, this inelastic property is robust over the entire population and various samples. Moreover, the price elasticity estimate obtained using Feldstein's specification, which includes *FG*, *RES*, and *GROW*, and that obtained without these variables are quite similar, 0.70 and 0.67. In addition, there is no statistical reason for selecting one model over the other when the corresponding $\bar{R}^2$ values for these nested models (0.81 and 0.79) are almost the same.

II. Cross-Section and Time-Series Results

The single cross-section results presented in the previous section are enriched by following a panel of 307 communities through time. Since considerable variation in matching rates exists across communities while much less variability is observed for the same municipality through time, a pooled cross-

[2] See his fn. 14, p. 81.

TABLE 3—DEMAND FOR LOCAL PUBLIC EDUCATION: 1970, 1973, 1975, 1976

Year	Sample Size	P	W	INC	SBG	$PRIV$	PUP	$\bar{R}^2$
A. Cross-Section Demand								
1970	341	−0.484	0.286	0.495	0.012	−0.044	−0.051	.53
		(2.70)	(11.33)	(8.79)	(2.44)	(3.12)	(1.05)	
1970	255	−0.471	0.262	0.706	0.009	−0.053	−0.116	.60
		(2.53)	(7.11)	(10.90)	(1.72)	(3.36)	(2.08)	
1970	143	−0.557	0.266	0.832	0.011	−0.081	−0.180	.74
		(2.72)	(5.16)	(11.89)	(2.10)	(4.20)	(3.37)	
1970	59	−0.722	0.524	0.477	0.013	−0.153	−0.087	.83
		(2.26)	(5.68)	(3.72)	(1.76)	(4.69)	(0.92)	
1973	333	−0.617	0.248	0.439	0.007	0.005	−0.271	.52
		(4.85)	(8.29)	(8.14)	(2.14)	(0.43)	(6.14)	
1973	259	−0.512	0.140	0.675	0.007	−0.013	−0.395	.59
		(3.58)	(3.42)	(10.01)	(1.82)	(1.04)	(7.77)	
1973	147	−0.749	0.266	0.712	0.012	−0.050	−0.411	.70
		(3.62)	(3.79)	(7.52)	(2.35)	(2.75)	(6.45)	
1973	58	−0.213	0.189	0.531	0.005	−0.047	−0.279	.66
		(0.83)	(1.89)	(3.67)	(0.78)	(2.04)	(3.00)	
1975	337	−0.642	0.185	0.444	0.014	−0.021	−0.316	.51
		(4.29)	(8.21)	(7.97)	(3.56)	(1.72)	(6.79)	
1975	260	−0.823	(0.249)	0.498	0.015	−0.018	−0.320	.59
		(5.08)	(7.29)	(7.58)	(3.77)	(1.43)	(6.04)	
1975	148	−0.468	0.117	0.661	0.009	−0.030	−0.340	.53
		(2.37)	(2.15)	(8.51)	(1.84)	(1.52)	(5.74)	
1975	60	−0.11	0.071	0.713	0.000	−0.045	−0.613	.67
		(0.39)	(0.83)	(5.64)	(0.08)	(1.51)	(6.71)	
1976	320	−0.435	0.184	0.411	0.003	0.007	−0.221	.56
		(3.80)	(10.20)	(9.58)	(0.91)	(0.73)	(5.61)	
1976	258	−0.398	0.160	0.454	0.002	0.008	−0.267	.56
		(2.96)	(6.05)	(8.71)	(0.72)	(0.74)	(6.10)	
1976	151	−0.331	0.112	0.540	0.007	0.013	−0.287	.54
		(2.03)	(2.51)	(8.09)	(1.77)	(0.73)	(5.47)	
1976	60	0.275	−0.062	0.540	−0.002	0.019	−0.540	.65
		(0.95)	(0.65)	(4.34)	(0.36)	(0.52)	(6.20)	
B. Pooled Cross-Section Time-Series Demand								
Pooled	−0.649	0.304	0.464	0.010	−0.001	−0.098		.47
OLS	(6.96)	(19.88)	(12.96)	(3.84)	(0.25)	(3.29)		
Fixed	−0.557	0.158	0.233	0.012	−0.114	−0.315		.20
Effects	(7.91)	(6.14)	(1.83)	(6.19)	(2.68)	(7.64)		

section and time-series model permits the disentanglement of the price and income effects from matching grants.

One approach to combining cross-section and time-series data is to write equation (1) as

$$\text{(2)} \quad \ln E_{it} = \beta_{0it} + \beta_1 \ln W_{it} + \beta_2 \ln P_{it} + \sum_{j=3}^{K} \beta_j \ln X_{jit} + \varepsilon_{it},$$

where the slope coefficients are assumed constant while the intercept varies over communities and time. For this paper a fixed effect, or dummy variable, model is adopted where it is assumed that $\beta_{0it} = \beta_0 + \mu_i + \lambda_t$, $\Sigma_i \mu_i = 0$, and $\Sigma_t \lambda_t = 0$. The community specific μ_is represent the influence of factors that vary among communities but not through time, and the time-specific λ_ts represent the influence of factors which vary with time but not across communities.[3]

[3]Alternatively, a random effect, or error components, model may be used to estimate (2) when the μ_i and λ_t

TABLE 4—OBSERVED MEASURES OF CATEGORICAL EQUITY IN LOCAL EDUCATION EXPENDITURE

$$\alpha_1 = \beta_w + \beta_p \gamma_{pw}$$

Year	Sample Size	α_1	β_w	β_p	γ_{pw}
1970	341	0.324	0.398	−0.484	0.153
1970	255	0.383	0.470	−0.471	0.185
1970	143	0.467	0.606	−0.557	0.250
1970	59	0.602	0.767	−0.722	0.228
1973	333	0.287	0.420	−0.617	0.215
1973	259	0.332	0.449	−0.512	0.229
1973	147	0.480	0.690	−0.749	0.281
1973	58	0.399	0.474	−0.213	0.351
1975	337	0.248	0.343	−0.642	0.148
1975	260	0.350	0.491	−0.823	0.171
1975	148	0.308	0.430	−0.468	0.261
1975	60	0.407	0.372	0.110	0.314
1976	320	0.199	0.261	−0.435	0.143
1976	256	0.240	0.302	−0.398	0.157
1976	151	0.260	0.351	−0.331	0.275
1976	60	0.296	0.186	0.275	0.400
Pooled Sample		0.216	0.316	−0.649	0.154

Additional observations on all Massachusetts cities and towns were obtained for 1973, 1975, and 1976 from which a panel of 307 communities has been formed.[4] All variables measured in dollars were deflated by the Consumer Price Index for Boston (1967 = 100). Table 3 presents both *OLS* results for various samples drawn from the four cross sections along with the estimates obtained from the pooled data.

Once again the price elasticity estimates are everywhere inelastic. Income is also everywhere inelastic although the cross-section results suggest that larger communities respond more to changes in income than smaller ones do.

III. Revised Policy Recommendation and Conclusion

The thrust of this comment is a challenge to Feldstein's argument: "The specific parameter estimates of the educational demand function imply that even complete wealth neutrality can be achieved...with matching grants that involve a relatively low elasticity of price with respect to wealth..." (p. 88). A measure of wealth neutrality, α_1, is related to the parameters of equation (1) as follows:

$$\alpha_1 = \beta_1 + \beta_2 \gamma_{pw} + \sum_{j=3}^{K} \beta_j \gamma_{x_{jw}}, \tag{3}$$

where γ_{pw} is the elasticity of price with respect to wealth and γ_{xjw} is the elasticity of X_j with respect to wealth. An adjusted wealth elasticity of expenditure can be defined

$$\beta_w = \beta_1 + \sum_{j=3}^{K} \beta_j \gamma_{x_{jw}} \tag{4}$$

vary randomly such that $E[\mu_i] = 0$ and $E[\lambda_t] = 0$. Using Jerry Hausman's specification test (1978) to choose among these models, the error components model was rejected in favor of the fixed effect model since I found evidence that the μ_i and λ_t were correlated with the X_{jit}. To accommodate the error components model's assumption about the error structure of λ_t, observations drawn in 1975 were omitted so that an equal time interval would separate each panel. In any event, the elasticities estimates were robust across these alternative models and included data panels.

[4]As explained above, where observed *PRIV* was nonpositive in any year, all observations on that community were deleted from the data.

thereby permitting the wealth neutrality measure to be rewritten as

$$\alpha_1 = \beta_w + \beta_p \gamma_{pw}. \tag{5}$$

The estimate of price-wealth elasticity (.15) obtained from pooling the four panels is quite similar to the low value reported by Feldstein (.18) for 105 communities in 1970. Using this along with his price elasticity estimate (−1.00) and adjusted wealth elasticity estimate (.37), he notes that the wealth-neutralizing grant scheme had reduced α_1 to "only .19" (p. 85). However, it is his relatively elastic estimate of β_p on which he concludes that the matching grant scheme could achieve (at a reasonable cost to the taxpayer) complete wealth neutrality. Using estimates obtained from the various sample sizes and years investigated here, Table 4 shows the calculation of the observed levels of categorical equity. Given that β_w and β_p are behavioral parameters, only γ_{pw} can be manipulated by policymakers to achieve categorical equity. Hence, the inelastic response to price movements observed here suggest that the grant scheme proposed by Feldstein could not effectively neutralize the influence of wealth in local education spending. Moreover, the large variations across communities in matching rates required by this policy suggest substantial price distortions that in turn may increase economic inefficiency.

REFERENCES

Barlow, Robin, "Efficiency Aspects of Local School Finance," *Journal of Political Economy*, September-October 1970, *78*, 1028–39.

Bergstrom, Theodore C., Rubinfeld, Daniel L. and Shapiro, Perry, "Micro-Based Estimates of Demand Functions for Local School Expenditures," *Econometrica*, September 1982, *50*, 1183–205.

Black, David E., Lewis, Kenneth A. and Link, Charles R., "Wealth Neutrality and the Demand for Education," *National Tax Journal*, June 1979, *32*, 157–64.

Bradford, David F. and Oates, Wallace E., "Suburban Exploitation of Central Cities," in Harold M. Hochman and George E. Peterson, eds., *Redistribution through Public Choice*, New York: Columbia University Press, 1974.

Brazer, Harvey E., "City Expenditures in the United States," Occasional Paper No. 66, National Bureau of Economic Research, 1959.

Feldstein, Martin S., "Wealth Neutrality and Local Choice in Public Education," *American Economic Review*, March 1975, *65*, 75–89.

Hausman, Jerry A., "Specification Tests in Econometrics," *Econometrica*, November 1978, *46*, 1251–71.

Inman, Robert P., "Micro-fiscal Planning in the Regional Economy: A General Equilibrium Approach," *Journal of Public Economics*, April 1977, *7*, 237–60.

Ladd, Helen F., "Local Education Expenditures, Fiscal Capacity, and the Composition of the Property Tax Base," *National Tax Journal*, June 1975, *22*, 145–58.

Lovell, Michael C., "Spending for Education: The Exercise of Public Choice," *Review of Economics and Statistics*, November 1978, *60*, 487–95.

Perkins, George M., "The Demand for Local Public Goods: Elasticities of Demand for Own Price, Cross Prices, and Income," *National Tax Journal*, December 1977, *30*, 411–22.

Public Education: Reply

By MARTIN FELDSTEIN*

To understand the significance of the price and wealth elasticities estimated in my 1975 article and reestimated for a variety of samples by George Perkins, it is useful to recall the legal controversy that surrounded education finance in the early 1970's. In the historic and much publicized case of *Serrano vs. Priest* (1971), the California Supreme Court held that California's system of educational finance violated California's state constitution because local educational outlays were related to local property values. Similar cases were being introduced in the courts of other states and in the U.S. Supreme Court.

Perhaps the most commonly discussed legal remedy in cases like *Serrano* was a plan referred to as "district power equalization" (*DPE*) by its original advocates, J. E. Coons, W. H. Clune, and S. D. Sugarman in their very influential book (1970). District power equalization is a matching grant formula that makes each percentage point of tax rate levied on the market value of local property produce the same revenue, independent of the actual local tax base.

To be more precise, let W_i be the tax base ("wealth") per pupil in school district i and θ_i be the tax rate chosen by school district i. The per pupil tax raised locally would therefore be $T_i = \theta_i W_i$. District power equalization would make the total per pupil revenue of district i proportional to θ_i but independent of W_i or $R_i = \theta_i W^*$, where W^* is the equivalent tax base implicitly assigned to all school districts by the *DPE* matching rate formula. The matching rate for district i (m_i) is the number of state level dollars given to district i for every dollar of revenue that they raise locally; therefore $1 + m_i = R_i/T_i = W^*/W_i$.

The *DPE* formula is important because it implies that the local "price" of educational outlays is proportional to wealth. If district i's price of buying educational outlays is defined as the amount that the district must produce in local tax revenue per dollar of total spending, the *DPE* formula implies $p_i = T_i/R_i = W_i/W^*$. In terms of the price-wealth elasticity of my earlier paper and Perkins' comment, district power equalization implies $\gamma_{pw} = 1$.

My reason for estimating price and wealth elasticities of demand for local school districts was to answer the following two questions. First, if the courts required the state governments to finance education in a way that eliminated the currently observed association between local wealth and educational outlays, what would be the appropriate matching formula? Second, if the courts mandated the district power equalizing rule, what would be the resulting association between wealth and educational outlays?

To make these ideas more precise, I decided to measure the association by the elasticity of per pupil education outlays with respect to per pupil wealth and called this elasticity the degree of wealth neutrality. By applying Theil's famous formula for specification bias, I showed that the wealth neutrality (α) can be written

$$(1) \qquad \alpha = \beta_w + \beta_p \gamma_{pw},$$

where β_w is the adjusted wealth elasticity defined in equation (4) of Perkins' comment, β_p is the price elasticity, and γ_{pw} is the elasticity of wealth with respect to price in Theil's auxiliary regression. Answers to the two questions follow immediately from equation (1).

First, complete wealth neutrality, defined as $\alpha = 0$, requires that γ_{pw} be set equal to $-\beta_w/\beta_p$. Perkins reports that, when all of the observations are pooled, the estimates

*Professor of Economics, Harvard University, Cambridge, MA 02138.

are $\beta_w = 0.316$ and $\beta_p = -0.649$, thereby implying that complete wealth neutrality can be achieved with $\gamma_{pw} = 0.49$. This is slightly less than half of the district power equalizing price-wealth elasticity of 1.0. It is in this sense that I wrote the conclusion quoted by Perkins: "The specific parameter estimates of the educational demand function imply that even complete wealth neutrality can be achieved... with matching grants that involve a relatively low elasticity of price with respect to wealth" (p. 88). My point was not to advocate such wealth-neutralizing matching grants, but rather to note that wealth neutrality, if required by the courts, could be achieved by a system of matching grants that implied a much lower price-wealth elasticity than the district power equalizing formula that had been discussed by the California court and that had been so widely advocated as "the" remedy to wealth disparities. I believe that Perkins' estimates therefore support my original conclusion.

Second, it is clear from equation (1) and Perkins' estimates that district power equalization would produce a negative wealth-expenditure elasticity. Since $\gamma_{pw} = 1$ implies $\alpha = \beta_w + \beta_p$, the pooled sample parameter estimates of $\beta_w = 0.316$ and $\beta_p = -0.649$ implies $\alpha = -0.333$. With *DPE*, the elasticity of educational outlays with respect to local wealth would be -0.333.

In recent years there has been much less discussion of litigation to achieve educational wealth neutrality. This may reflect an increased understanding that there is substantial heterogeneity of income and wealth within school districts, making the correlation between educational outlays and income or wealth much lower at the individual level than at the more aggregate level of school district averages. It may also reflect a greater public acceptance of the legitimacy and desirability of local responsibility for educational decisions.

REFERENCES

Coons, J. E., Clune, W. H. and Sugarman, S. D., "*Private Wealth and Public Education*," Cambridge, 1970.

Feldstein, Martin, "Wealth Neutrality and Local Choice in Public Education," *American Economic Review*, March 1975, *65*, 75–89.

Perkins, George M., "Public Education: Comment," *American Economic Review*, September 1984, *74*, 814–19.

Serrano vs. Priest, L. A. 29820, Superior Court No. 938254, cited in *Harvard Education Review*, November 1971, *41*, 503.

Economics Departmental Rankings: Comment

By Barry T. Hirsch, Randall Austin, John Brooks, and J. Bradley Moore*

In a recent issue of the *Review* (1982), Philip Graves, James Marchand, and Randall Thompson (G-M-T) provide a ranking of economics departments in the United States based on page counts of articles published during 1974–78 in twenty-four leading journals. The purpose of this comment is to update the G-M-T rankings, based on publications from 1978 through spring 1983. We also provide rankings for the top forty economics departments outside the United States. Finally, we compare concentration ratios measuring the dispersion in published pages among U.S. departments between the 1974–78 and 1978–83 periods.

In order to ensure comparability, we rank departments using the same method as did G-M-T. Of course, no simple methodology can capture accurately the many dimensions which comprise research quality. Rankings are based on *AER*-standardized page counts in the twenty-four leading journals used by G-M-T.[1] School assignment is based on the author's designation in the journal at the time of publication. Where articles have multiple authors, each is accorded a weight equal to $1/n$, where n is the number of authors. The rankings measure standardized pages in economics journals by all persons affiliated with each school, rather than by members of economics departments alone. Thus, universities with active research faculties in business and finance, statistics, agricultural economics, and law will tend to be more highly ranked. For this reason, we do not present rankings which weight by economics department size.[2] Special issues of journals are included only when article quality is judged to be similar to those in a regular issue.[3]

Table 1 presents U.S. departmental rankings based on the total number of *AER*-standardized pages from 1978 to 1983 (spring) for the top 240 colleges and universities. In addition, we include for each school its ranking based on unstandardized total pages, the G-M-T 1974–78 ranking, and the existence of a Ph.D. program in economics.[4] We leave interpretation of these results to the reader.

Table 2 presents rankings of the top forty non-U.S. economics departments, including their worldwide ranking based on data from Tables 1 and 2. Rankings for non-U.S. schools, of course, are particularly sensitive to the choice of journals. The London School of Economics easily ranks first (and fourth worldwide), while the next six universities have only small differences in total pages.

In Table 3, we present concentration ratios for the leading 1, 5, 10, 25, 50, and 100 U.S.

*Hirsch is Professor of Economics and Austin, Brooks, and Moore are graduate assistants, University of North Carolina, Greensboro, NC 27412–5001. The study was conducted as part of a university review of graduate programs.

[1]The twenty-four journals are given in G-M-T, fn. 3, p. 1132. The time period includes 1978–82, plus all 1983 issues prior to June. The *AER*-standardized page weights were provided by Graves. Most were close to but below one; thus, for most schools the sum of standardized pages is less than the total of unstandardized pages. The largest weight is for *JASA* (this affects measurably the rankings for a few schools, most notably Iowa State), while the smallest weight is for *Oxford Economic Papers*.

[2]In addition, faculty sizes are not readily available for many schools. Wyn Owen and Larry Cross provide faculty sizes for universities with graduate programs in economics. These figures, combined with those presented in Table 1, allow the interested reader to rank departments by pages per faculty member.

[3]For example, *JPE* supplements are included since articles there typically are original contributions similar in length and quality to those in regular issues. By contrast, the *AER Proceedings* issues are excluded since papers there are generally shorter nontechnical papers summarizing research published elsewhere or in progress. It is not clear how G-M-T treated special issues.

[4]Information on Ph.D. programs is provided by Owen and Cross. Auburn University has begun a Ph.D program since the *Guide* was published.

TABLE 1—U.S. ECONOMICS DEPARTMENTAL RANKINGS BASED ON *AER*-STANDARDIZED PAGES

	1978–83		1974–78
	Total Pages[a]	Rank[b]	G-M-T Rank
1. *Chicago	2976.1	1	1
2. *Harvard	2427.4	2	2
3. *Stanford	1996.1	3	3
4. *Pennsylvania	1660.0	4	5
5. *Yale	1502.9	5	7
6. *Northwestern	1461.6	6	11
7. *MIT	1442.2	7	6
8. *Wis-Madison	1386.0	8	4
9. *UC-Berkeley	1281.2	10	9
10. *UCLA	1246.3	11	8
11. *Cornell	1236.6	12	21
12. *Columbia	1229.1	9	17
13. *Princeton	1186.6	13	10
14. *Minnesota	1123.5	14	20
15. *Michigan	1062.4	15	12
16. *Rochester	983.5	16	14
17. *Ill-Urbana	941.0	19	15
18. *New York	901.3	18	18
19. *Carnegie-Mellon	900.8	21	28
20. *Washington	877.6	17	13
21. *Ohio State	850.6	20	19
22. *UNC-Chapel Hill	713.5	23	16
23. *UC-San Diego	659.0	24	33
24. *Purdue	644.1	25	23
25. *Virginia	620.0	22	22
26. *Texas A&M	565.8	30	30
27. *Rutgers	556.4	26	35
28. *VPI	549.2	32	26
29. *UC-Santa Barbara	526.6	29	61
30. *Penn State	526.3	33	25
31. *Michigan State	524.5	31	27
32. *Florida	521.1	35	29
33. *USC	518.9	27	37
34. *UC-Davis	517.7	28	58
35. *Maryland	481.9	34	24
36. *Indiana	458.8	37	42
37. *Iowa	445.7	39	39
38. *Cal Tech	404.7	36	71
39. *Johns Hopkins	388.5	38	41
40. *Duke	372.4	40	40
41. *SUNY-Buffalo	369.5	43	45
42. *Auburn	368.9	41	88
43. *NC State	350.0	46	83
44. *Arizona	336.3	49	87
45. Delaware	323.3	45	52
46. *Vanderbilt	321.2	44	48
47. *Boston	318.4	42	66
48. *Houston	314.5	47	38
49. *Iowa State	312.5	63	43
50. *Wis-Milwaukee	309.0	53	53
51. *Georgia	303.4	50	34
52. *SUNY-Stonybrook	302.7	48	47
53. UNC-Greensboro	294.3	52	128
54. *Texas	283.6	60	31
55. *Brown	279.8	51	32
56. *South Carolina	275.9	59	72
57. *SMU	270.6	57	44
58. Dartmouth	269.6	56	63
59. *Boston College	259.1	55	46
60. *Missouri	256.0	64	56
61. *Arizona State	253.6	61	54
62. *Ill-Chicago Circle	252.6	54	55
63. *Kansas	251.6	62	68
64. *Wayne State	244.1	58	59
65. Texas-Dallas	218.6	70	173
66. *Utah	217.2	66	69
67. *Tennessee	216.4	67	107
68. Miami (Ohio)	211.5	65	—
69. *Temple	211.1	71	74
70. *Syracuse	203.4	73	62
71. *Florida State	203.2	81	57
72. *Washington State	200.1	68	60
73. *Georgia State	199.2	77	64
74. *Oklahoma	188.4	75	86
75. *SUNY-Binghamton	186.6	69	51
76. *Oregon	185.6	74	65
77. *Pittsburgh	173.8	80	67
78. Brigham Young	173.2	72	182
79. *Kentucky	168.5	78	78
80. *George Washington	168.5	86	75
81. *Tulane	166.3	79	73
82. *SUNY-Albany	164.0	85	91
83. Wesleyan	157.7	76	92
84. *Wyoming	146.8	87	84
85. *Washington-St. Louis	144.0	83	81
86. *Colorado	142.5	90	89
87. *New Mexico	139.3	82	80
88. *S. Ill.-Carbondale	138.9	84	100
89. *Connecticut	138.5	92	97
90. *Georgetown	131.4	89	85
91. George Mason	130.9	88	113
92. Clemson	126.0	99	125
93. *N. Illinois	124.5	93	112
94. *Hawaii	120.1	96	70
95. CUNY-Baruch	119.6	98	—
96. *Rice	116.8	105	49
97. *Massachusetts	116.0	95	50
98. *Colorado State	115.5	100	108
99. Illinois State	111.2	97	110
100. *LSU	107.0	101	82
101. *Williams	104.5	94	138
102. *Oklahoma State	103.6	104	101
103. Brandeis	103.2	91	—
104. *Cincinnati	98.9	102	93
105. *Lehigh	91.6	106	115
106. *Tufts	88.2	108	103
107. San Diego State	87.0	103	—
108. Emory	82.2	107	153
109. *West Virginia	78.1	110	109
110. Va. Commonwealth	76.4	113	160
111. Bowdoin	76.1	111	132
112. Miami (Florida)	72.1	114	—
113. San Jose State	71.9	117	139
114. *Montana State	65.8	120	209
115. Mass-Boston	63.3	115	102
116. *Texas Tech	62.9	127	120
117. *American	62.6	109	79
118. *CUNY-Grad. Sch. & U. Ctr.	62.4	112	—
119. Oregon State	62.0	121	135
120. CUNY-Queens Coll.	59.1	125	—
121. *Notre-Dame	58.2	118	157
122. Cal State-Long Beach	57.4	123	198
123. Amherst	56.6	116	95
124. Denver	56.0	122	—
125. SUNY-Brockport	54.5	138	180
126. *Fordham	53.7	126	183
127. Memphis State	51.7	147	176
128. Marquette	50.5	128	134
129. Missouri-Rolla	50.5	157	155
130. Maine	50.4	135	175
131. Cleveland State	50.2	132	121
132. Loyola	49.5	129	225
133. Georgia Tech	49.4	119	96
134. *Case Western	48.9	139	90
135. Minn-St. Paul	48.1	170	—
136. Missouri-St. Louis	47.9	140	137
137. *Kansas State	47.5	156	167
138. Santa Clara	45.9	131	122
139. UC-Santa Cruz	45.6	133	123
140. William and Mary	44.8	134	168
141. Bradley	44.7	150	—
142. Indiana (PA)	44.4	136	124
143. Wis-Parkside	42.2	137	174
144. S. Ill-Edwardsville	41.4	141	—
145. Holy Cross	40.8	146	165
146. Swarthmore	40.7	148	77

(Continued)

TABLE 1—CONTINUED

		1978–83		1974–78			1978–83		1974–78
		Total pages[a]	Rank[b]	G-M-T Rank			Total pages[a]	Rank[b]	G-M-T Rank
147.	Smith College	39.8	124	–	196.	Clarkson	18.4	197	111
148.	Cal State-Northridge	39.8	143	104	197.	St. Joseph's (PA)	18.4	195	–
149.	S. Florida	39.1	152	145	198.	*New Hampshire	17.7	203	150
150.	*UC-Riverside	39.0	182	94	199.	Tampa	17.5	211	–
151.	Missouri-KC	38.5	130	140	200.	Maryland-Baltimore	17.1	214	–
152.	Union College	37.4	149	141	201.	Kent State	17.0	213	159
153.	*UC-Irvine	37.4	145	151	202.	Ill.-Wesleyan	16.9	198	–
154.	*Claremont	37.2	142	154	203.	Carleton College	16.8	221	144
155.	*Nebraska	36.7	144	105	204.	Southern Conn State	16.7	202	–
156.	*Clark	36.5	158	127	205.	*Bryn Mawr	16.6	181	193
157.	Cal State-Chico	36.2	154	202	206.	Pomana College	16.5	210	–
158.	New Orleans	34.9	153	152	207.	SUNY-Geneseo	16.5	199	114
159.	Texas-San Antonio	34.8	174	–	208.	Hampden-Sydney	16.3	259	–
160.	Rhode Island	34.6	167	117	209.	San Francisco	16.1	217	–
161.	Cal State-Fullerton	33.6	163	161	210.	E. Michigan	15.8	204	217
162.	Old Dominion	33.2	168	186	211.	Bowling Green	15.7	212	143
163.	*Arkansas	32.9	166	–	212.	Montana	15.6	205	149
164.	Tulsa	31.7	164	133	213.	Western Michigan	15.5	260	178
165.	Naval Post Grad	31.1	187	–	214.	Gordon College	15.3	224	–
166.	Ball State	29.6	159	–	215.	Trinity-San Antonio	15.3	206	–
167.	Wellesley	29.4	161	188	216.	Monmouth College	15.1	215	–
168.	Cal Poly. State	29.2	180	231	217.	US Naval Academy	14.9	207	–
169.	DePaul	29.1	165	–	218.	Lewis & Clark	14.9	219	–
170.	Hamilton	29.1	162	126	219.	Tenn-Chattanooga	14.6	229	–
171.	East Carolina	28.8	176	169	220.	Barnard	14.4	200	–
172.	*Alabama	28.8	169	106	221.	Clarion State	14.4	196	–
173.	Appalachian State	28.6	178	–	222.	Middlebury	14.2	230	–
174.	*Mid. Tennessee State	28.4	171	–	223.	Bentley	14.1	231	235
175.	*Alabama State	28.1	151	–	224.	W. Ill	14.1	208	223
176.	Vermont	27.8	189	116	225.	Baylor	13.8	227	–
177.	Shippensburg State	27.8	177	–	226.	Mississippi State	13.7	270	136
178.	Texas-Arlington	27.7	194	221	227.	St. Cloud State	13.7	233	236
179.	New Mexico State	26.1	181	–	228.	Wis-Lacrosse	13.3	225	–
180.	Sangamon State	26.0	172	–	229.	Chapman	13.1	222	228
181.	Nevada-Reno	25.5	175	99	230.	SU-New Orleans	12.8	226	–
182.	Cal State-LA	24.9	155	119	231.	W. Kentucky	12.4	218	–
183.	Cal State-Hayward	24.6	160	190	232.	Wabash	12.3	236	–
184.	Conn-Southeastern	24.6	173	–	233.	Lowell	12.2	228	185
185.	*Northeastern	23.4	190	98	234.	Fairfield	12.2	237	–
186.	Weber State	22.7	188	–	235.	Colby College	12.1	223	197
187.	Cal State-Sacramento	22.2	191	–	236.	Lawrence	12.0	238	203
188.	John Carroll	21.9	193	207	237.	Oberlin	12.0	239	131
189.	Virginia State	21.8	185	–	238.	*Mississippi	12.0	244	162
190.	Oakland (MI)	21.8	201	–	239.	*Ohio University	11.8	216	158
191.	James Madison	21.5	186	191	240a.	C.W. Post College	11.3	247	–
192.	*Catholic	20.9	184	195	240b.	South Alabama	11.3	248	–
193.	Central Michigan	19.4	179	–	240c.	CUNY-City College	11.3	245	–
194.	*Utah State	19.3	220	196	240d.	Louisiana Tech	11.3	249	–
195.	Radcliffe	19.0	192	–					

Note: Asterisk indicates Ph.D. program; a dash denotes unranked by G-M-T for 1974–78; a double dash denotes ranking by G-M-T not clear.
[a]Total *AER*-standardized pages.
[b]Rank by unstandardized total pages.

departments for each of the two time periods. These ratios are calculated by

$$CR(m) = \sum_{i=1}^{m} PAGES_i / \sum_{i=1}^{240} PAGES_i,$$

where *PAGES* are total *AER*-standardized pages, i indexes school by rank, and m represents the number of leading schools. Note that *CR* is not a true concentration ratio since the denominator measures the number of pages by the top 240 U.S. schools and not the total pages in the journals for each period (a number which is not available from the G-M-T study).

Publication in economics is highly concentrated among the leading universities. While a number of departments have moved significantly up or down in the rankings between 1974–78 and 1978–83, the aggregate concentration ratios of published pages have

TABLE 2—NON-U.S. ECONOMICS DEPARTMENTAL RANKINGS BASED ON *AER*-STANDARDIZED PAGES, 1978–83

	Total Pages[a]	Non-U.S. Rank[b]	Worldwide Rank (1)	Worldwide Rank (2)
1. London Sch. of Econ.	1878.0	1	4	4
2. British Columbia	961.8	3	18	19
3. Western Ontario	833.4	4	24	20
4. Hebrew (Jerusalem)	830.1	5	25	26
5. Oxford	817.4	2	26	16
6. Tel-Aviv	816.7	7	27	28
7. Toronto	814.5	6	28	27
8. Australian National	653.9	8	31	29
9. Cambridge	527.5	9	37	35
10. Queen's (Ontario)	514.8	10	44	40
11. Warwick	306.4	11	61	58
12. Bristol	280.3	12	66	61
13. Carleton	264.0	13	71	66
14. York (U.K.)	249.6	15	77	74
15. Essex	246.7	17	78	79
16. Birmingham	237.0	14	80	70
17. McMaster	236.7	18	81	82
18. Southampton	222.5	20	82	89
19. Bonn	213.3	16	86	77
20. New South Wales	195.9	19	93	87
21. Birkbeck (London)	187.8	22	95	94
22. Monash	185.3	21	98	90
23. Alberta	180.6	24	99	98
24. Simon Fraser	174.8	23	100	95
25. Manchester	166.2	27	106	104
26. L'Ecole Polytechnique	159.6	25	108	101
27. Newcastle-upon-Tyne	157.1	28	110	109
28. Kyoto	154.8	31	111	116
29. U. College of London	149.9	26	112	102
30. McGill	148.6	29	113	110
31. Tokyo	138.7	30	119	112
32. Norwegian Sch. Econ. & Bus. Ad.	134.1	35	121	122
33. Catholique de Louvain	129.6	33	124	120
34. Canterbury (N.Z.)	121.8	37	127	126
35. Stockholm	115.8	32	132	118
36. Bar-Ilan	113.5	38	134	130
37. Paris	111.3	36	135	123
38. Ben Gurion	111.1	42	137	142
39. Montreal	109.6	43	138	144
40. Leeds	102.5	47	143	152

Note: Worldwide Rank: Col. (1) by *AER*-standardized pages; Col. (2) by unstandardized total pages.

[a,b]See Table 1.

TABLE 3—CONCENTRATION OF PUBLISHED PAGES AMONG LEADING SCHOOLS[a]

	1974–78	1978–83
$CR(1)$	.053	.053
$CR(5)$	.205	.187
$CR(10)$	.321	.308
$CR(25)$	.550	.554
$CR(50)$	.735	.743
$CR(100)$	.910	.912

[a]$CR(m)$ is defined in text.

remained virtually unchanged between the periods. It is widely believed that research activity over time has increased relatively more at middle- and lower-ranked schools. While this may be correct, the constancy of CR over the two periods suggests that relatively little of the research by faculty at such schools has been published in the leading economics journals.

REFERENCES

Graves, Philip E., Marchand, James R. and Thompson, Randall, "Economics Departmental Rankings: Research Incentives, Constraints, and Efficiency," *American Economic Review,* December 1982, *72,* 1131–41.

Owen, Wyn F. and Cross, Larry R., *Guide to Graduate Study in Economics and Agricultural Economics,* 6th ed., Boulder: Economics Institute, 1982.

Economics Departmental Rankings: Comment

By TIMOTHY D. HOGAN*

Over the past decade, several articles have studied the relative publishing performance of economists associated with the economics departments of U.S. universities (John Siegfried, 1972; William Moore, 1973; Albert Niemi, 1975; V. Kerry Smith and Steven Gold, 1976; Philip Graves, James Marchand, and Randall Thompson, 1982, for example). In these studies, credit for publications has generally been assigned to the institution listed by the author on the article itself. In addition to their "football rankings" aspect, most of these studies have interpreted data on publication output in terms of measures of relative quality of the faculties/graduate programs. For some purposes, such compilations based upon listed affiliation may be appropriate (as a long-term measure of quality, or to measure relative productivity at that institution over the study period, for example). But if we are interested in evaluating the quality of current economics faculties/graduate programs, this methodology is flawed. Since academic economists (prolific publishers in particular) are a mobile group, significant changes may occur over a relatively short period of time and not be reflected by publication data collected in terms of this "byline" method. Additional bias is also introduced by the fact that some authors listing affiliation at a given institution may have been students rather than faculty (or faculty/staff affiliated with other agencies within the institution and not associated with the economics program).

This paper presents an alternate method of quantifying the publication performance of academic departments and demonstrates that this alternative procedure produces significantly different results when compared with the usual method. The paper also discusses how these data can be interpreted as measures of interinstitution flows of publishing faculty. Examination of these data in this light provides evidence of systematic patterns of such human capital flows over the 1960–79 period.

I. The Alternate Publication Measure

This study has compiled data on the volume of journal publication during the 1970–79 period for the *current faculties* of the economics department at those U.S. institutions offering graduate degrees in economics. That is, rather than tabulating publications on the basis of listed affiliation, each publication was associated with the author's name (authors' names), and credit for publication given only when an author was listed as a member of the economics department in the *Guide to Graduate Study in Economics, Agricultural Economics, and Related Fields* (supplemented by college/university catalogues where necessary).[1] Data on the number of pages published in the four top-rated economics journals (*American Economic Review*, *Econometrica*, *Journal of Political Economy*, and *Quarterly Journal of Economics*) by the current faculties of the fifty economics departments whose publications in those four journals totaled at least 100 pages over the 1970–79 period is presented in column A of Table 1.[2] The same publication

*Department of Economics, Arizona State University, Tempe, AZ 85287. I gratefully acknowledge the comments and suggestions of two anonymous referees and of my colleagues, John McDowell and Robert D. St. Louis, at earlier stages of this analysis.

[1] Only publications authored by persons listed either as a faculty member of the department of economics, or as a member of another department but explicitly affiliated with the economics department in a joint capacity, were included in the compilations.

[2] Contributions were defined to include articles, review articles, notes, communications, comments, and memorials. Replies, rejoinders, or corrections by an original author and book reviews were not included. When a contribution was written by more than one author, the pages were divided equally. The data for the other three journals were adjusted for differences in

data have also been aggregated on the basis of listed affiliation to produce total pages figures for the journals for the 1970–79 period, but using compilation procedures equivalent to that used by Graves, Marchand, and Thompson (hereafter, G-M-T) and most of the other analyses. These *AER*-equivalent page totals for the same fifty institutions are set forth in column B of Table 1. Rankings (on the basis of all U.S. institutions with graduate economics departments) based upon these publications data also are included in the table. For comparison, publication rankings for the institutions as compiled by G-M-T are presented in column E in Table 1.

While Harvard is found to rank first by either compilation method, comparison of the relative publishing performance based alternatively on current faculties versus listed affiliation demonstrates considerable differences in rankings among the fifty institutions (columns C and D). Among the top-rated departments, for example, Princeton ranks much higher on a current faculty basis, while the University of Chicago and the University of California-Berkeley departments rank substantially lower than they did for publication totals based upon listed affiliation. A similar pattern of differences is found when rankings based upon "current affiliation" data and the G-M-T data are compared. In addition to those divergences in relative ratings mentioned above, institutions such as Minnesota, MIT, Yale, and Brown are found to rank substantially higher on the basis of publications of current faculty than they did in the G-M-T study, while certain other departments, such as Illinois and North Carolina fell in rankings.

As a formal test of the hypothesis that the tabulation of data on publishing performance based upon current faculties produces different results than does compilation on the basis of listed affiliation, the Wilcoxon matched-pairs test was employed using the publication figures for the fifty institutions presented in Table 1. In addition, since one potential source of variation between the current affiliation data set and the publications data compiled by G-M-T might be differences in both the time period and sample of journals from which the two data sets were collected (four top journals for the 1970–79 period vs. twenty-four journals over the 1974–78 period), the same test was also performed to test for significant differences between data compiled by this study on the basis of listed affiliation and the G-M-T data for the same fifty institutions. These statistical tests show no statistically significant difference between the rankings produced from either of the data sets compiled on the basis of listed affiliation. Comparing my current affiliation and listed affiliation data sets, on the other hand, the null hypothesis of no significant difference in rankings was rejected by the Wilcoxon procedure.[3] Thus, the evi-

page volume and aggregated to produce combined data reported in terms of *AER*-page equivalents. The four journals selected for this sample were those found to rank either one through four or at least at the very top of the quality rankings in several studies that attempted to rank economics journals (see, for example, R. G. Hawkins, L. S. Ritter, and I. Walter, 1973, or M. C. Lovell, 1973). It would, of course, have been preferable to expand the sample, but due to the amount of effort involved in compilation of the data, the analysis was limited to these four prestigious journals. Given that the limited sample includes journals published by Harvard and Chicago, the possible existence of "home journal" bias was investigated. When alternative total page counts were compiled in which all contributions to the *QJE* were excluded, Harvard was still ranked first. On the basis of page counts excluding the *JPE*, however, Chicago slipped to fourteenth compared with fifth on the basis of page totals for all four journals. Since the editorship of the *AER* was associated with Brown for much of the period of analysis, a similar alternative page count was compiled excluding *AER* contributions. With those data, the relative publication output measured for Brown also declined with its current faculty output ranking dropping to twentieth (rather than fourteenth based upon total pages in all four journals). For a detailed analysis of the effects of "in-house" editorships, see John McDowell and Ryan Amacher, 1983.

[3]Since the total number of pages credited to authors at each institution differed between the two data sets (20,869 pages for the "listed-at-time-of-publication" data set vs. 15,380 pages for the "current faculties" data), the current affiliation totals in column A were actually compared with a series of "expected" numbers of pages based upon the proportion of the total listed affiliation contributed by each institution (computed from column B). The computed z statistic for the Wilcoxon test (with $N = 50$) was -2.129—statistically significant at the 0.03 level for a two-tailed test. Similarly, the listed affiliation

(Continued on p. 831)

TABLE 1—U.S. ECONOMICS DEPARTMENTS WITH CURRENT FACULTIES CONTRIBUTING MORE THAN 100 TOTAL PAGES IN TOP FOUR JOURNALS, 1970–79[a]

Department	Total Pages by Current Faculty (A)	Total Pages by Listed Affiliation (B)	Current Faculty Ranking (C)	Listed Affiliation Ranking (D)	Ranking per G-M-T (E)
Harvard	1022	1982	1	1	2
Princeton	808	798	2	6	10
MIT	780	945	3	4	6
Yale	754	1049	4	3	7
Chicago	712	1820	5	2	1
Stanford	661	891	6	5	3
Wis-Madison	568	730	7	9	4
Minnesota	552	695	8	11	20
UCLA	525	698	9	10	8
Penn	506	688	10	12	5
Northwestern	496	735	11	8	11
Columbia	486	644	12	13	17
UC-Berkeley	424	789	13	7	9
Brown	408	407	14	19	32
Washington	401	471	15	17	13
Michigan	362	616	16	14	12
Rochester	338	567	17	15	14
Virginia	327	381	18	20	22
Cornell	299	421	19	18	21
SUNY-Buffalo	293	263	20	24	45
Maryland-College Park	278	267	21	23	24
Ohio State	211	308	22	22	19
VPI	211	225	22	28	26
UC-San Diego	189	249	24	25	33
New York	189	341	24	21	18
Boston	184	52	26	71	65
Texas A&M	178	229	27	27	30
SUNY-Stony Brook	177	81	28	57	47
Arizona	175	60	29	66	84
Michigan State	174	202	30	32	27
Ill-Urbana	171	212	31	30	15
George Washington	162	87	32	55	74
UC-Santa Barbara	157	164	33	35	61
Carnegie-Mellon	154	559	34	16	28
UNC-Chapel Hill	154	247	34	26	16
Duke	153	137	36	40	40
Florida State	152	143	37	39	57
Houston	151	135	38	41	38
USC	147	113	39	45	37
Johns Hopkins	144	204	40	31	41
Boston College	129	180	41	33	46
Ill-Chicago Circle	124	111	42	47	55
Mass-Amherst	123	213	43	29	49
SUNY-Binghamton	114	96	44	52	51
Indiana	113	116	45	42	42
Cal Tech	113	115	45	44	70
SMU	112	168	47	34	44
UC-Davis	107	87	48	55	58
NC State	107	113	48	45	81
Florida	105	65	50	64	29

[a]Top four journals: *American Economic Review*; *Econometrica*; *Journal of Political Economy*; *Quarterly Journal of Economics*.

TABLE 2—DIFFERENCES IN RANKING ON THE BASIS OF CURRENT FACULTY VS. LISTED-AT-TIME-OF-PUBLICATION AFFILIATION

	1970–79[a]			1960–72[b]		
	Affiliation		Difference	Affiliation		Difference
	Current (A)	Listed (B)	in Ranking (C)	Current (D)	Listed (E)	in Ranking (F)
Harvard	1	1	–	1	2	+1
Princeton	2	6	+4	7	8	+1
MIT	3	4	+1	3	4	+1
Yale	4	3	−1	4	3	−1
Chicago	5	2	−3	2	1	−1
Stanford	6	5	−1	6	5	−1
Wis-Madison	7	9	+2	9	10	+1
Minnesota	8	11	+3	8	11	+3
UCLA	9	10	+1	13	18	+5
Penn	10	12	+2	5	7	+2
Northwestern	11	8	−3	11	13	+2
Columbia	12	13	+1	18	14	−4
UC-Berkeley	13	7	−6	16	6	−10
Brown	14	19	+5	12	19	+7
Washington	15	17	+2	19	15	−4
Michigan	16	14	−2	15	12	−3
Rochester	17	15	−2	10	16	+6
Virginia	18	20	+2	25	22	−3
Cornell	19	18	−1	17	17	–
SUNY-Buffalo	20	24	+4	29	25	−4
Maryland	21	23	+2	22	33	+11
Ohio State	22	22	–	27	30	+3
VPI	22	27	+5	32	36	+4
New York	24	21	−3	31	27	−4
Texas A&M	25	26	+1	29	39	+10
Michigan State	26	30	+4	21	20	−1
Ill.-Urbana	27	28	+1	23	21	+2
Carnegie-Mellon	27	16	−12	14	9	−5
UNC-Chapel Hill	28	25	−3	28	32	+4
Duke	30	33	+3	23	28	+5
Johns Hopkins	31	29	−2	20	22	+2
Indiana	32	34	+2	35	26	−9
NC State	33	36	+3	41	35	−6
Vanderbilt	34	42	+8	26	31	+5
Washington	34	37	+3	42	34	−8
Iowa State	36	38	+2	33	29	−4
Kansas	37	34	−3	39	37	−2
Pittsburgh	38	39	+1	45	43	−2
Purdue	38	31	−7	33	24	−9
Penn State	40	32	−8	43	40	−3
Rice	41	40	−1	36	38	+2
Iowa	42	41	−1	40	42	+2
Claremont	43	44	+1	44	45	+1
Texas	44	43	−1	38	44	+6
Oregon	45	45	–	37	41	+4

[a]*Source:* Computed from Table 1.
[b]*Source:* My earlier paper, Table 1.

dence implies that the differences in rankings between my data set compiled by current affiliation and the G-M-T data are primarily the result of the alternative compilation procedure, rather than due to differences in journal coverage and/or time period.[4]

To examine changes in relative publishing performance from the early 1970's to the end of the decade, Table 2 presents the relative rankings with respect to the relative publication output in the same four journals during the 1960–72 period of the forty-five top-rated (as of 1969) Ph.D. programs in economics together with the rankings data for the 1970–79 period compiled by this study.[5] Comparison of the affiliation-at-end-of-period data for the two periods (column A vs. D) provide some interesting insights. Nine of the departments whose current faculties were among the ten most productive over the 1970–79 period also had similar results with respect to their 1973 faculties, but the relative positions within this elite group were subject to considerable shifting—for example, Chicago slipping from second to fifth and Princeton jumping from seventh to second. Substantial differences in relative publication records between 1973 and current faculties are also evident for departments below the top ten ranks. Most striking is the fact that twelve of the departments ranked among the top forty-five Ph.D. programs in the 1969 A.C.E. study are not found among the group of fifty departments with current faculties having top publication records over the 1970–79 period.[6]

II. Indirect Evidence of Human Capital Flows

The differences in these alternate measures of publishing output can also be viewed in a more general way. Such shifts in rankings are an indication of the flows of human capital that are continually occurring within the acedemic community. It is primarily through research activity and publications that academicians gain both professional reputations and financial rewards. Several examples were identified among the fifty-school sample in which departments ranked substantially higher on the basis of publishing output of their current faculties than was the case for a publishing index based upon listed affiliation at time of publication. It appears that such institutions offered sufficient pecuniary (or nonpecuniary) benefits to attract economists

data compiled by this study (from four top journals for the 1970–79 period) were compared with the G-M-T data set for the same fifty schools. The computed z statistic was -0.642—not statistically different from zero ($p = 0.52$ for a two-tailed test).

[4]Using the listed affiliation of the author for compiling publication performance data can result in crediting publications of authors who were actually faculty members in other departments (business school or law school faculties, for example) or nonfaculty researchers not directly associated with the graduate economics program at those institutions, or who were undergraduate or graduate students at the time of publication (see G-M-T, p. 1132). To examine this issue, several issues of the annual *National Faculty Directory* were consulted to identify the departmental affiliation of authors at time of publication for several of the high-ranking schools. From this single source, it was possible to confirm that a substantial portion of the total publications credited to departments were, in fact, authored by noneconomics faculty, students, etc. For example, 35 percent of the University of Chicago total was identified as authored by individuals not on the economics faculty at the time of publication, and the same process identified several other institutions for which 20 percent or more of the credited pages were authored by noneconomics faculty, students, or nonfaculty researchers at those schools—Northwestern (28 percent), University of Washington (22 percent), University of California-Berkeley (21 percent), Stanford (20 percent), and MIT (20 percent). The proportion of publications authored by noneconomics faculty was not found to be consistent across institutions; in the case of UCLA, for example, only 8 percent of the total pages were so identified. If these noneconomics faculty located in the business school, law school, etc., are actively involved in the economics Ph.D. program, such exclusions may not be warranted. In the compilation of the current faculty data, however, faculty with joint appointments or otherwise considered to be associated with the graduate economics program by the institution itself as represented in the *Guide to Graduate Study in Economics* were included.

[5]These ranks are based upon publication data for the 1960–1972 period compiled as part of my previous study (1975). The sample of forty-five schools was based upon ratings by the 1969 American Council of Education study (K. D. Roose and C. J. Anderson, 1970).

[6]These institutions were Claremont, Iowa, Iowa State, Kansas, Oregon, Pennsylvania State, Pittsburgh, Purdue, Rice, Texas, Vanderbilt, and Washington University (St. Louis).

who had elsewhere demonstrated ability to produce publishable research. These institutions were able to improve the overall quality of their faculties ("expand the 'pool' of current expertise," in the terminology used by G-M-T, p. 1132) at the expense of other departments by entering the market and bidding these high-quality human resources away from their existing institutions.

Comparison of the changes in rankings with compilation based upon current vs. listed affiliation for two different time periods provides insight into possible longer-term patterns in such flows of publishing economists among the set of graduate economics faculties. Differences in relative rankings between "end-of-period" (current) affiliation and "time-of-publication" (listed) affiliation for the 1960–72 period for the forty-five departments included in my earlier study (1975), and for the 1970–79 period for those same schools are presented in Table 2. Among the top five schools, both Princeton and MIT ranked relatively higher on the basis of current faculty publications (both 1973 and 1979) rather than listed affiliation, while Yale and Chicago are found to have lower relative rankings for both time periods. Among the other schools in the sample, departments such as Minnesota, Brown, VPI, and Vanderbilt are also identified as "net-gainers" during both time periods. Other departments, such as University of California-Berkeley, Carnegie-Mellon, and Purdue, are found to have consistently lower current faculty vis-à-vis listed affiliation rankings. Evidence of such long-term flows over the 1960–72 and 1970–79 time period is consistent with a pattern in which certain departments have tended to act as sites where relatively inexperienced economists gain skills and/or are identified (through publications) as possessing high-quality human capital, and these successful publishers are then hired away by other institutions who systematically pursue a strategy of allowing the "capital formation" and/or screening process to occur at other schools.

Examination of the variances of the differences in rankings between current and listed affiliates for the two time periods indicates that such differences were greater in the earlier period than changes occurring during the 1970–79 period. It appears that, while stable patterns in these shifts seem to exist over both time periods, the absolute volume of such movement declined during the 1970's as the boom period for higher education of the 1960's faded away.[7]

[7]An F-test of the ratio of the variances of the two series indicates that variances of the differences in ranks for the earlier period is significantly greater than that for the later period testing at the 0.05 significance level:

$$F=\frac{S^2(1960\text{–}1972)}{S^2(1970\text{–}1979)}=\frac{22.604}{13.084}=1.728$$

with 44, 44 degrees of freedom. Alternative hypotheses might be offered to explain this phenomenon. Two reasonable (and related) explanations might be the following: 1) a lessening in mobility of established faculty due to changing labor market conditions—previous studies have analyzed the shifts in supply and demand for faculty in the academic labor market during the 1960's and 1970's (Allan Cartter, 1976; Richard Freeman, 1979; and Charles Scott, 1979), and Cartter has identified sharp declines in interinstitutional mobility of established Ph.D.s that accompanied these market shifts (see pp. 153–62); 2) a decline in the phenomenon of "rich" departments trying to enhance their reputations quickly by buying established, high-quality faculty—during higher education's golden years of the 1960's, many institutions found themselves in the advantageous situation of receiving big budget increases and/or allocation of additional faculty slots, along with encouragement to pursue the goals of expanding their graduate programs and improving the quality and reputations of their departments. In many cases, however, the monies available and the support for such activities dried up with the end of higher education enrollment growth, the decline in federal funding, and leaner state budgets during the 1970's.

REFERENCES

Cartter, Allen C., *Ph.D.'s and the Academic Labor Market*, New York, 1976.

Freeman, Richard B., "Demand for Labor in a Non-Profit Market: University Faculty," in D. S. Hamermesh, ed., *Labor in the Public and Non-Profit Sectors*, Princeton: Princeton University Press, 1979.

Graves, Philip E., Marchand, James R. and Thompson, Randall, "Economics Departmental Rankings: Research Incentives,

Constraints, and Efficiency," *American Economic Review*, December 1982, *72*, 1131–41.

Hawkins, R. G., Ritter, L. S. and Walter, I., "What Economists Think of Their Journals," *Journal of Political Economy*, July/August 1973, *81*, 1017–32.

Hogan, Timothy D., "Statistical Measures of Faculty Quality: Correcting Previous Bias," in *Proceedings* of the American Statistical Association, Business and Economics Section, 1975, 415–19.

Lovell, M. C., "The Production of Economic Knowledge," *Journal of Economic Literature*, March 1973, *11*, 27–55.

McDowell, John M. and Amacher, Ryan C., "Economic Value of an In-House Editorship," Economics Faculty Working Papers, EC 83/84-3, Arizona State University, 1983.

Moore, William J., "The Relative Quality of Graduate Programs in Economics, 1958–1972: Who Published and Who Perished," *Western Economic Journal*, July 1973, *11*, 1–23.

Niemi, Albert W., Jr., "Journal Publication Performance During 1970–1974: The Relative Output of Southern Economics Departments," *Southern Economic Journal*, July 1975, *42*, 97–106.

Owen, Wyn F. and Cross, Larry R., *Guide to Graduate Study in Economics and Agricultural Economics*, 6th ed., Homewood: Richard D. Irwin, 1982.

Roose, K. D. and Anderson, C. J., *A Rating of Graduate Programs*, Washington: American Council of Education, 1970.

Scott, Charles E., "The Market for Ph.D. Economists: The Academic Sector," *American Economic Review Proceedings*, May 1979, *69*, 137–42.

Siegfried, John J., "The Publishing of Economic Papers and its Impact on Graduate Faculty Ratings, 1960–1969," *Journal of Economic Literature*, March 1972, *10*, 31–49.

Smith, V. Kerry and Gold, Steven, "Alternative Views of Journal Publication Performance During 1968–71 and 1970–74," *Eastern Economics Journal*, April 1976, 109–13.

Gale Research Company, *The National Faculty Directory*, Detroit, Michigan, 1971–78.

Economics Departmental Rankings: Reply and Errata

By Philip Graves, James Marchand, and Randall Thompson*

While authors often wonder whether they are read, the preceding comments by Barry Hirsch et al. and Timothy Hogan, and Paul Davis and Gustav Papanek's article (1984), as well as other submissions and correspondence to the managing editor of this *Review* from individuals, leave no doubt that our previous rankings were closely scrutinized. The following remarks address certain criticisms of our earlier paper, acknowledge some mistakes, and point to some inadequacies in efforts to rank departments.

The comments differ according to time period or methodology employed, yet the findings are "broadly" similar to our prior findings. This claim is limited to the results for the more productive departments and Ph.D.-granting institutions. The general agreement is exemplified by Hogan's finding using the Wilcoxon matched-pairs technique of no significant difference between our rankings and his.

Davis and Papanek appear most critical of rankings of our type or the Hirsch et al. update. They made a number of valid points, using a different methodology and covering a later time period (1978 and 1981) than our study. Their rankings show relative upward movement for certain expanding institutions when compared to our rankings. However, our rankings exhibit a similar pattern of relative upward movement when compared to the earlier rankings by Albert Niemi (1975).

While Hogan and Davis-Papanek offer a good discussion and qualification of their ranking methodologies relative to ours and Hirsch et al., one further qualification relative to proper interpretation remains. Both the Hogan and Davis-Papanek ranking methodologies look to the past more than do our and the Hirsch et al. approaches. Their ranking methodology might be superior or more indicative of current department and research quality and quantity if research quality and quantity depends largely or only on accumulated human capital, assuming no depreciation. However, if other factors, such as salary incentives, secretaries, research assistance, teaching load, research and/or teaching assistance, age of the faculty, etc. matter, then the question is, how well are faculty doing in their current employment? Our approach and that of Hirsch et al. reflects this more accurately than the alternatives which are more oriented towards the past. The regressions presented in our original article indicate that perhaps some of these other factors do matter although causation is difficult to ascertain in the reduced-form model presented there.

Some of our correspondence (and some of the papers on this subject rejected by the managing editor of this *Review*) attributed strong downward bias and injustice in our ranking technique for departments or universities with "unusual" but quality faculty research in public policy, agricultural economics, finance, econometrics and statistics, law and economics, etc. Footnotes 1 and 6 in our earlier article acknowledge the potential for biases of this nature. However, it does not seem that such biases would be large, particularly for the top schools. Our list of journals included general journals (*AER*, *JPE*, *QJE*, *SEJ*, and the like) open to quality research from many fields. Furthermore, surveyed journals included at least some specialist journals serving as outlets for research in many areas. We explored the difference made to the rankings in unreported earlier work which reduced the number of journals and which dropped "home" journal output from the respective school's total. The results changed little for the top schools. One has to draw the line somewhere and we selected the

*Economics Department, University of Colorado; Economics Department, University of Mississippi; Treasury Department, Atlantic Richfield Company, Los Angeles, respectively.

ERRATA TABLE—1974–78 DATA

Changes		Additions		Deletions
A. Table 1—*AER*-Equivalent-Sized Pages...				
Baruch College[a]	132.3	U Toledo[b]	10.13	SUNY-Oswego[c]
Miami (Ohio)[d]	129.98			
Miami	20.23			
*N. Illinois[h]	67.83			
B. Table 2—Pages per Economics Department Faculty Member...				
Baruch College[a]	5.75	U Toledo[b]	.84	SUNY-Oswego[c]
Miami (Ohio)[d]	6.50	Occidental[e]	4.08	
*N. Illinois[h]	2.83	Indiana (PA)[f]	4.54	
		Wright State[g]	3.10	

Note: Asterisk indicates Ph.D. program.
[a]As reported by Albert Zucker, Chairman, Economics Department, Baruch College.
[b]As reported by Edward Shapiro, Economics Department, University of Toledo.
[c]As reported by James Cicarelli, former Chairman, SUNY-Oswego.
[d]As reported by James Dunlevy and Robert Newman, Economics Department, Miami University.
[e]As reported by James Halstead, Economics Department, Occidental College.
[f]As reported by Donald Walker, Economics Department, Indiana University of Pennsylvania.
[g]As reported by John Blair, Chairman, Economics Department, Wright State University.
[h]As reported by Prem Laumas, Chairman, Economics Department, Northern Illinois University.

Niemi journals to ease intertemporal comparisons, a useful tradition continued in the Hirsch et al. contribution.

In conclusion, the preceding comments and the Davis-Papanek paper are a welcome addition to our work and should give the profession a solid, though inevitably imperfect, picture of the quality of economics departments as measured by their research output.

It is now appropriate to turn to the errors and inadequacies of our earlier paper. These mistakes are errors of attribution and omission. Much correspondence concerning potential error resulted from our failure to clearly point out that we followed Niemi in methodology. This meant that for articles with multiple authors, pages were split between institutions. Lack of clarity regarding this led many to suspect underattribution.

The correct attribution of published work to faculties is made difficult by the rather incomplete descriptions of affiliations used by some authors and by some of the twenty-four journals examined. It is worth noting that rankings of this sort would be facilitated by a complete, specific description of author(s) affiliation. It is also important to note that inadequacies in affiliation description must as well complicate and expose to error the development of citation indexes and almost all ranking methodologies. In our case, the usual problems of attribution were complicated through the expansion of the number of colleges and universities examined. Having developed a count of pages by institution, developing a faculty count proved more difficult than anticipated. To develop the faculty number, we use Wyn Owen and Larry Cross' *Guide*, our sample survey, and departmental listings in the college and university catalogs of the University of Colorado library. Other attempts at faculty counting were deemed too costly given our resource base. As a consequence, our original Table 2 only includes colleges and universities for which the number of faculty was available through the above efforts and did not cover all institutions in Table 1 which would otherwise have appeared in Table 2. It was intended that a clarifying footnote be added to Table 2 to eliminate confusion. Our errors of attribution and omission produced

correspondence from various members of the profession. The Errata Table presents those corrections to the original Tables 1 and 2 which we have uncovered as a result.

The two most detrimental errors were our failure to adequately count the publications of Baruch College in the City University of New York system and those at Miami University, Oxford, Ohio. This led to errors in both Tables 1 and 2. The remaining reported errors in Table 2 have been corrected through the thoughtful correspondence from current or former department members of those institutions.

REFERENCES

Davis, Paul and Papanek, Gustav, "Faculty Ratings of Major Economic Departments by Citations," *American Economic Review*, March 1984, *74*, 225–30.

Graves, Philip E. et al. "Economics Departmental Rankings: Research Incentives Constraints and Efficiency," *American Economic Review*, December 1982, *72*, 1131–41.

Hirsch, Barry et al., "Economics Departmental Rankings: Comment," *American Economic Review*, September 1984, *74*, 822–26.

Hogan, Timothy D., "Economics Departmental Rankings: Comment," *American Economic Review*, September 1984, *74*, 827–33.

Niemi, Albert W., Jr., "Journal Publication Performance During 1970–1974: The Relative Output of Southern Economics Departments," *Southern Economic Journal*, July 1975, *41*, 97–106.

Owen, Wyn F. and Cross, Larry R., *Guide to Graduate Study in Economics and Agricultural Economics*, 4th ed., Homewood: Richard D. Irwin, 1977.

Preliminary Announcement of the Program

NINETY-SEVENTH ANNUAL MEETING OF THE AMERICAN ECONOMIC ASSOCIATION

Dallas, Texas, December 27–30, 1984

Thursday, December 27, 1984

10:00 A.M. EXECUTIVE COMMITTEE MEETING

Friday, December 28, 1984

8:00 A.M. LABOR CONTRACTS AND MACROECONOMIC PERFORMANCE
Presiding: JOSEPH E. STIGLITZ, Princeton University
Papers: COSTAS AZARIADIS, University of Pennsylvania, AND RUSSELL COOPER, Yale University
Nominal Wage-Price Rigidity as Rational Expectations Equilibria
DANIEL J. B. MITCHELL, University of California-Los Angeles
Wage Flexibility in the United States: Lessons from the Past
MARTIN L. WEITZMAN, Massachusetts Institute of Technology
Macroeconomic Properties of Profit-Sharing Contracts
Discussants: GUILLERMO CALVO, Columbia University
JO ANNA GRAY, Washington State University

8:00 A.M. RATIONALITY, UNCERTAINTY AND INTERACTION
Presiding: GORDON C. WINSTON, Williams College
Papers: ROGER A. MCCAIN, Fordham University
Adjustment Costs, Approximate Optima and Imperfect Rationality
CHARLES F. MASON AND FREDERICK F. STERBENZ, University of Wyoming
Uncertain Product Quality, Advertising and Intermediary Certification
SHLOMO MAITAL, Technion, Haifa
How People Interact: Toward a General Theory of Externalities
Discussants: PAUL ALBANESE, Middlebury College
RICHARD LANGLOIS, University of Connecticut

8:00 A.M. ECONOMICS OF THE PLASTIC ARTS
Presiding: J. CARTER MURPHY, Southern Methodist University
Papers: J. M. MONTIAS, Yale University
The Market for Art in Seventeenth-Century Netherlands
ROGER W. WEISS, University of Chicago
The International Art Market and Economic Development
WILLIAM D. GRAMPP, University of Illinois-Chicago
The Uneasy Case of Museum Subsidies
Discussants: JAN DEVRIES, University of California-Berkeley
STANISLAW WELLISZ, Columbia University

8:00 A.M. CREDIT AND ECONOMIC INSTABILITY
Presiding: JAMES S. EARLEY, University of California-Riverside
Papers: BENJAMIN M. FRIEDMAN, Harvard University
(To be announced)
ROBERT N. POLLIN, University of California-Riverside
(To be announced)
ALBERT M. WOJNILOWER, First Boston Corporation
(To be announced)
Discussants: HYMAN P. MINSKY, Washington University-St. Louis
JAMES S. EARLEY, University of California-Riverside

8:00 A.M. THE IMPACT OF EXCHANGE RATES
Presiding: CARLOS J. DIAZ-ALEJANDRO, Columbia University
Papers: SUSAN RANNEY, University of Washington
Devaluation and Microeconomic Incentive for International Migration
ANN HELWEGE, Tufts University
The Appropriateness of Fiscal Stimulants in the Context of a Foreign Exchange Crisis
PEKKA AHTIALA, University of Tampere, Finland
Expectations, the Supply Side, and Fiscal and Monetary Policies under Flexible Exchange Rates
Discussants: CARLOS J. DIAZ-ALEJANDRO, Columbia University
JONATHAN EATON, Yale University

8:00 A.M. HEALTH INSURANCE (Joint Session with the Health Economics Research Organization)
Presiding: CHARLES E. PHELPS, University of Rochester
Papers: JODY L. SINDELAR, University of Chicago
The Market Structure for Health Insurance: Empirical Estimates
FRANK SLOAN, Vanderbilt University, AND ROSS MULLNER, American Hospital Association
Uncompensated Hospital Care: Who Gets and Who Pays
WILLIAM G. ANLYAN, JR., University of Virginia, AND JOSEPH LIPSCOMB, Duke University
The National Health Care Trust Plan: Blueprint for Inflation Control and Long-Term Care Provision
Discussants: DAVID KASS, Federal Trade Commission
CATHERINE J. ELLIOTT, College of William and Mary
CHARLES E. PHELPS, University of Rochester

8:00 A.M. ECONOMIC IMPLICATIONS OF MARITAL INSTABILITY FOR CONSUMER DECISION MAKING AND EXPENDITURE
Presiding: SIMONE CLEMHOUT, Cornell University
Papers: MURRAY BROWN, State University of New York-Buffalo
On the Existence and Properties of Household Bargaining Models
ROBERT A. POLLAK, University of Pennsylvania
A Transactions Costs Approach to Allocation and Distribution within the Family
SIMONE CLEMHOUT AND JANET MCCLAIN, Cornell University
The Influence of Intra-Household Interactive Patterns on Consumer Decisions and Expenditures
Discussants: MARILYN MANSER, U.S. Bureau of Labor Statistics
MARJORIE MCELROY, Duke University

8:00 A.M. ISSUES IN INTERNATIONAL PRODUCTIVITY COMPARISONS
Presiding: J. R. NORSWORTHY, U.S. Bureau of the Census
Papers: ZVI GRILICHES, Harvard University
International Productivity Comparisons: Some Empirical Puzzles
W. ERWIN DIEWERT, University of British Columbia
On the Use of Exchange Rates in International Productivity Comparisons
CATHERINE J. MORRISON, Tufts University
Capacity Utilization and Productivity Growth: A U.S.–Japan Comparison
Discussants: CHARLES R. HULTEN, The Urban Institute
MIEKO NISHIMIZU, World Bank

8:00 A.M. ISSUES IN AGRICULTURAL ECONOMICS
Presiding: LANCE TAYLOR, Massachusetts Institute of Technology
Papers: JOHN STRAUSS, Yale University
Does Better Nutrition Raise Farm Productivity?
MINE E. CINAR AND T. J. MURPHY, Loyola University of Chicago
An Analysis of Short-Run Market Efficiency in U.S. Grain Markets
J. J. JORDAN, R. L. SHEWFELT, S. E. PRUSSIA, AND W. C. HURST, University of Georgia
Using an Hedonic Price Function to Evaluate Postharvest Technologies
Discussants: JERE BEHRMAN, University of Pennsylvania
PHILIP ABBOTT, Purdue University

8:00 A.M. ECONOMIC AGING IN BRITAIN AND THE UNITED STATES
Presiding: R. D. NORTON, Mount Holyoke College
Papers: MANCUR OLSON, JR., University of Maryland
The Unviability of Laissez-Faire in Britain and America

CAROL HEIM, University of Massachusetts-Amherst
Decline and Renewal in Britain and the United States: The Role of Underdeveloped Areas within Mature Economies
BRUCE NORTON, Wellesley College
Marxian Theories of Stagnation and Long Waves: A Review
Discussants: WILLIAM N. PARKER, Yale University
PETER LEWIN, University of Texas-Dallas

8:00 A.M. CHINESE ECONOMIC DEVELOPMENT AND TRADE (Joint Session with the Committee on Asian Economic Studies)
Presiding: JAMES T. H. TSAO, U.S. International Trade Commission and George Mason University
Papers: JAMES T. H. TSAO, U.S. International Trade Commission and George Mason University
Chinese Development Strategy and Technology Importation
EUGENE K. LAWSON, Deputy Assistant Secretary of Commerce, East Asia and the Pacific
U.S.-China Trade: Behind the Myth
ALBERT KEIDEL, Wharton Econometric Forecasting Associates
China's Post-Mao Economy: A New Dynasty Begins
Discussants: PETER H. LINDERT, University of California-Davis
ANTHONY M. TANG, Vanderbilt University

8:00 A.M. NEW PERSPECTIVES ON H. H. GOSSEN, *The Laws of Human Relations* (1854) (Joint Session with the History of Economics Society)
Presiding: RUDOLPH C. BLITZ, Vanderbilt University
Papers: DON L. COURSEY, University of Wyoming
Gossen and Hierarchical Preferences
NICHOLAS GEORGESCU-ROEGEN, Vanderbilt University
Time and Scarcity in Gossen's Economic System
JOHN K. WHITAKER, University of Virginia
Gossen as a Pioneer of General Equilibrium Analysis
Discussants: KENNETH J. ARROW, Stanford University
KLAUS HENNINGS, Technical University of Hannover, FRG
DONALD MCCLOSKEY, University of Iowa

10:15 A.M. REGIONAL CONFLICT AND ECONOMIC POLICY
Presiding: JACK E. ADAMS, University of Arkansas-Little Rock
Papers: CHARLES W. HOWE, University of Colorado
Project Benefits and Costs from National and Regional Accounting Stances: Methodology and Case Study
NILES M. HANSEN, University of Texas-Austin
Regional Consequences of Structural Changes in the National and International Division of Labor
RICHARD J. CEBULA, Emory University
Regional Consequences of Changing Population Distribution and Factor Proportions
Discussants: ROGER BURFORD, Louisiana State University
RONALD MOOMAW, Oklahoma State University

10:15 A.M. TOPICS IN LAW AND ECONOMICS
Presiding: RICHARD O. ZERBE, University of Washington
Papers: MARILYN J. SIMON, U.S. Department of Justice
Protect Quarterly and the Allegation of Legal Cost
DAVID L. SHAPIRO, San Francisco State University
Public Power Policy in the Pacific Northwest: The Legal Fall Out
RICHARD O. ZERBE, University of Washington
Ethical Basis for Decisions in Law and Economics

10:15 A.M. FRONTIERS IN DEMOGRAPHIC ECONOMICS
Presiding: T. PAUL SCHULTZ, Yale University
Papers: PAUL A. SAMUELSON, Massachusetts Institute of Technology
How Biologic and Economic Approaches Differ and Agree in Demographic Analysis
ODED STARK AND DAVID BLOOM, Harvard University
The New Economics of Labor Migration
JAMES J. HECKMAN, University of Chicago, AND V. JOSEPH HOTZ, Carnegie-Mellon University
Fertility Behavior over the Life Cycle: A Survey of Recent Findings

Discussants: JOHN G. RILEY, University of California-Los Angeles
WARREN C. SANDERSON, State University of New York-Stony Brook

10:15 A.M. COMPARATIVE INTERNATIONAL STUDIES OF THE ROLE OF NONPROFIT ORGANIZATIONS
Presiding: EGON NEUBERGER, State University of New York-Stony Brook
Papers: AVNER BEN-NER, University of Haifa
A Life Cycle Model of Nonprofit Organizations: Birth and Demise
ESTELLE JAMES, State University of New York-Stony Brook
The Private Nonprofit Provision of Education in Comparative Perspective
BURTON WEISBROD, University of Wisconsin-Madison
A Model of Donations to Nonprofit Organizations
Discussant: MARK SCHLESINGER, Harvard University

10:15 A.M. TOPICS IN THE MULTINATIONAL CORPORATION
Presiding: DAVID MCCLAIN, Boston University
Papers: JEAN-FRANCOIS HENNART, University of Pennsylvania
Vertical Integration in the World Aluminum and Tin Industries: A Comparative Study of the Choice Between Market and Intrafirm Coordination
J. M. VISENE, Eramus University, Rotterdam
Macroeconomic Adjustment under Foreign Direct Investment
KRISTIN HALLBERG, Colby College
Exchange Risk and the Intercountry Distribution of Foreign Direct Investment
Discussants: BRUCE KOGUT, Massachusetts Institute of Technology
STEPHEN P. MAGEE, University of Texas-Austin

10:15 A.M. THE PACIFIC CHALLENGE FOR WORLD ECONOMIC LEADERSHIP
Presiding: LAWRENCE KRAUSE, The Brookings Institution
Papers: STAFFAN B. LINDER, Stockholm School of Economics
Pacific Protagonist: Implications of the Rising Role of the Pacific
WALT W. ROSTOW, St. Anthony College, Oxford
Is There a Need for Leadership: Japanese or American?
CARL KAYSEN, Massachusetts Institute of Technology
Can the United States Rise to the Pacific Challenge?
Discussants: CARL E. BEIGIE, Dominion Securities, Ames
LAWRENCE KRAUSE, The Brookings Institution

10:15 A.M. TRADE AND ADJUSTMENT POLICIES IN LATIN AMERICA
Presiding: DANIEL M. SCHYDLOWSKY, Boston University
Papers: MOSHÉ SYRQUIN, Harvard University
Adjustment Processes to External Shocks: A Comparative Analysis
DANIEL M. SCHYDLOWSKY, Boston University
Macroeconomic Effects of Nontraditional Exports in Peru
SIMÓN TEITEL, Inter-American Development Bank and Catholic University of America, AND F. THOUMI, Inter-American Development Bank and American University
From Import Substitution to Exports: The Recent Experience of Latin American Manufactured Exports
Discussants: MARIO BLEJER, International Monetary Fund
V. CORBO, Universidad Catolica de Chile
J. DE MELO, World Bank

10:15 A.M. RISK PERCEPTION AND MARKET PERFORMANCE
Presiding: MANCUR OLSON, University of Maryland
Papers: HERMAN B. LEONARD AND RICHARD ZECKHAUSER, Harvard University
Does the Market Recognize Financial Risk?
W. KIP VISCUSI, Duke University
Are Individuals Bayesian Decision Makers?
HOWARD KUNREUTHER, University of Pennsylvania, AND ROBIN HOGARTH, University of Chicago
Does Risk Ambiguity Affect the Performance of Insurance Markets?
Discussants: GEORGE AKERLOF, University of California-Berkeley
KENNETH J. ARROW, Stanford University
GARY S. BECKER, University of Chicago

10:15 A.M. TRENDS IN GOVERNMENT FINANCE
Presiding: ANN F. FRIEDLAENDER, Massachusetts Institute of Technology
Papers: THOMAS M. HOLLOWAY AND JOSEPH C. WAKEFIELD, U.S. Department of Commerce
Sources of Changes in the Size and Composition of the Federal Budget Deficit
HAROLD G. VATTER AND JOHN F. WALKER, Portland State University
Real Public Sector Employment, Growth, Wagner's Law, and Economic Growth in the United States
CARL S. SHOUP, Columbia University
Implications of Jointness Over Users for Recent Analysis of Growth in Government Share
Discussants: LORAINE EDEN, Brock University
ROBERT W. KILPATRICK, U.S. Office of Management and Budget

10:15 A.M. UNDERSTANDING CAPITALISM: APPRAISING DEVELOPMENTS IN RADICAL POLITICAL ECONOMY
Presiding: LEONARD W. RAPPING, University of Massachusetts-Amherst
Papers: SAMUEL BOWLES AND RICHARD EDWARDS, University of Massachusetts-Amherst
Radical Economics—Re-Statement
ROBERT M. SOLOW, Massachusetts Institute of Technology
Radical Economics: Reappraisal I
JOSEPH E. STIGLITZ, Princeton University
Radical Economics: Reappraisal II
Discussants: EILEEN APPLEBAUM, Temple University
ARTHUR MCEWAN, University of Massachusetts-Boston

10:15 A.M. TOPICS IN LOW-INCOME WORK AND ETHNICITY
Presiding: THOMAS D. BOSTON, Atlanta University
Papers: DAVID N. LABAND, University of Maryland, AND BERNARD F. LENTZ, Ursinus College
Age-Earning Profiles of Self-Employed Proprietors
REBECCA M. BLANK, Princeton University
Analyzing Trickle Down: How Labor Market Opportunities Among the Poor are Affected by General Economic Growth
WALTER S. MCMANUS, University of Florida
Earnings of Ethnic American Men: The Role of Language Skills
Discussants: (To be announced)
THOMAS D. BOSTON, Atlanta University

10:15 A.M. MALTHUS: 150 YEARS AFTER HIS DEATH (Joint Session with the History of Economics Society)
Presiding: BETTE H. POLKINGHORN, California State University-Sacramento
Papers: JOHN P. HENDERSON, Michigan State University
The Malthus-Ricardo Relationship
WILLIAM O. THWEATT, Vanderbilt University
Malthus's Library
SALIM RASHID, University of Illinois-Urbana
Malthus's Misuse of Demographic Principles
Discussants: WILLIAM D. GRAMPP, University of Illinois-Chicago
GEOFFREY H. GILBERT, Hobart and William Smith Colleges
DAVID M. LEVY, George Mason University

10:15 A.M. PEACE ECONOMICS I: DEFENSE EXPENDITURES, TRADING RELATIONS, AND FOREIGN AID POLICIES IN THE WORLD ECONOMY (Joint Session with the Peace Science Society (International))
Presiding: WALTER ISARD, Cornell University
Papers: LAWRENCE R. KLEIN, University of Pennsylvania
Impacts of Alternative Military Cutbacks and Foreign Aid Programs
SOLOMON POLACHEK, State University of New York-Binghamton
Dyadic Dispute: An Economic Approach to International Conflict Using Trade Data

12:30 P.M. AEA/AFA JOINT LUNCHEON
Presiding: CHARLES P. KINDLEBERGER, Massachusetts Institute of Technology and Brandeis University
Speaker: ALEXANDRE LAMFALUSSY, Bank for International Settlements
How to Live with Our Present International Monetary Arrangements

2:30 P.M. THE THEORY OF ECONOMIC ORGANIZATIONS
Presiding: MARTIN L. WEITZMAN, Massachusetts Institute of Technology

Papers: R. K. SHAH, Hoover Institution, AND JOSEPH E. STIGLITZ, Princeton University
The Architecture of Economic Systems: Hierarchies and Polyarchies
JIM MARCH, Stanford University
Adaptive Organizations
KENNETH J. ARROW, Stanford University
Informational Structure of the Firm
Discussants: SANFORD J. GROSSMAN, University of Chicago
PAUL MILGROM, Yale University

2:30 P.M. OPEN AND SEALED-BID AUCTIONS
Presiding: JOHN G. RILEY, University of California-Los Angeles
Papers: ERIC MASKIN, Massachusetts Institute of Technology
Recent Advances in Auction Theory
ROBERT G. HANSEN, Dartmouth College
Empirical Testing of the Revenue Equivalence Theorem
VERNON L. SMITH, University of Arizona
Experimental Auctions
Discussant: WILLIAM F. SAMUELSON, Boston University

2:30 P.M. THE FUNCTION OF THE FIRM IN THE SOVIET ECONOMIC SYSTEMS
Presiding: JOHN MOORE, Hoover Institution
Papers: STEVEN ROSEFIELDE AND R. W. PFOUTS, University of North Carolina-Chapel Hill
Market Socialism: The Noncompetitive Soviet Solution
JOHN BONIN, Wesleyan University
Innovation and the Soviet Firm
HOLLAND HUNTER, Haverford College
Enterprise Microplanning: Excessive Tautness
Discussants: MARTIN SPECHLER, University of Iowa
EGON NEUBERGER, State University of New York-Stony Brook

2:30 P.M. MACROECONOMIC ANALYSIS OF LEADING INTERWAR AUTHORITIES
Presiding: ANDREW F. BRIMMER, Brimmer & Co., Inc.
Papers: L. DWIGHT ISRAELSEN, Utah State University
Marriner S. Eccles, Chairman of the Federal Reserve Board
WILLIAM A. DARITY, JR., University of North Carolina-Chapel Hill
Rudolf Hilferding, Finance Minister in the Weimar Republic
DICK NANTO, Library of Congress
Korekiyo Takahashi, Prime Minister of Japan and Finance Minister, 1918–22 and 1931–36
Discussants: RAYMOND W. GOLDSMITH, Yale University
TAKAFUSA NAKAMURA, University of Tokyo

2:30 P.M. PERSPECTIVE ON THE EXTERNAL DEBT SITUATION
Presiding: BAREND A. DE VRIES, World Bank
Papers: WILLIAM R. CLINE, Institute for International Economics
External Debt and the International System: A Reassessment
MARIO HENRIQUE SIMONSEN, Getulio Vargas Foundation
Adjustment in a Major Debtor Economy
EDUARDO WIESNER, International Monetary Fund
Excessive Indebtedness: Lessons of the Past, Options for the Future
Discussants: JEFFREY FRANKEL, University of California-Berkeley
BASIL G. KAVALSKY, World Bank
HENRY WALLICH, Board of Governors of the Federal Reserve System

2:30 P.M. ISSUES IN MACROECONOMICS: A SESSION HONORING OTTO ECKSTEIN (Joint Session with the Eastern Economic Association)
Presiding: DALE W. JORGENSON, Harvard University
Papers: HENDRIK S. HOUTHAKKER, Harvard University
The International Agenda
ALICE M. RIVLIN, The Brookings Institution
Thoughts on Public Expenditure
ROBERT M. SOLOW, Massachusetts Institute of Technology
Reflections on Macroeconomic Model Building: Confessions of a DRI Addict

Discussants: ROGER BRINNER, Data Resources Inc.
HERBERT STEIN, American Enterprise Institute

2:30 P.M. PROBLEMS OF INEFFICIENCY IN THE SOVIET ECONOMY: THE CAUSES AND POSSIBLE SOLUTIONS (Joint Session with the Association for Comparative Economic Studies)
Presiding: ELIZABETH CLAYTON, University of Missouri-St. Louis
Papers: FYODOR I. KUSHNIRSKY, Temple University
The Soviet Economy at the Crossroads: Search for Efficiency
VLADIMIR KONTOROVICH, University of Pennsylvania
Technological Progress and Productivity Slowdown in the Soviet Economy, 1951–82
HELEN H. M. BOSS, Harvard University
Mind Over Matter in the Adoption of the Material Product System in the USSR? The Gerschenkron-Keyes Controversy Revisited
Discussants: ALICE GORLAND, Oakland University
ROGER SKURSKI, University of Notre Dame

2:30 P.M. DISCRETION VS. PRECOMMITMENTS IN MONETARY POLICY
Presiding: ROBERT HODRICK, Northwestern University
Papers: ALEX CUKIERMAN, Carnegie-Mellon University and Tel Aviv University, AND ALLAN H. MELTZER, Carnegie-Mellon University
Discretion vs. Precommitments—Some Positive and Normative Issues
DAVID GORDON, University of Rochester
Time Inconsistency, Rational Expectations, and Dynamic Equilibrium
STANLEY W. BLACK AND MICHAEL K. SALEMI, University of North Carolina-Chapel Hill
Government Policy and the Risk Premium in Foreign Exchange Markets
Discussants: HERSCHEL I. GROSSMAN, Brown University
ROBERT FLOOD, Northwestern University

2:30 P.M. THE FUTURE OF ECONOMIC HISTORY: ITS SOURCE MATERIALS
Presiding: CAROLYN SHAW BELL, Wellesley College
Papers: JANET NORWOOD, Bureau of Labor Statistics
Labor Statistics in a Changing World: 1884; 1984; 2084
GEOFFREY MOORE, Columbia University
Improving the Data Base for Current and Future Economic Analysis
CHARLES RENFRO, C. G. Renfro & Associates
Economic Measurement: On the Making of Bricks Without Straw
Discussants: ANNE CARTER, Brandeis University
JOSEPH DUNCAN, Dun & Bradstreet Corporation

2:30 P.M. EQUITY CONSIDERATIONS IN PUBLIC POLICY ANALYSIS
Presiding: RONALD E. GRIESON, University of California-Santa Cruz
Papers: KARL E. CASE, Wellesley College, AND MICHAEL MCPHERSON, Williams College
The Analysis of Subsidy Taxes in the U.S. Distributional System
GORDON TULLOCK, George Mason University
For Equitable Income Distributions
SANDRA BAUM, Wellesley College, AND SAUL SCHWARTZ, Tufts University
Equity, Envy and Higher Education
Discussants: THOMAS F. POGUE, University of Iowa
RONALD E. GRIESON, University of California-Santa Cruz

2:30 P.M. EMPIRICAL ANALYSIS OF INDUSTRIAL MARKETS
Presiding: TIMOTHY F. BRESNAHAN, Stanford University
Papers: HENRY W. CHAPPELL, JR. AND RONALD P. WILDER, University of South Carolina
R&D, Firm Size, and Concentration: Evidence from the FTC Line of Business Survey
DAVID I. ROSENBAUM, University of Wisconsin-Madison
The Titanium Dioxide Industry: An Empirical Analysis of Form Limit Pricing
RICHARD N. CLARK, U.S. Department of Justice
SICs as Delineators of Economic Markets

2:30 P.M. BLACK AND WHITE WOMEN IN POVERTY
Presiding: LAURIE BASSI, Georgetown University
Papers: EMILY P. HOFFMAN, Western Michigan University
The Feminization of Poverty: Economic and Demographic Determinants

JULIANNE MALVEAUX, San Francisco State University, AND BERNADETTE CHACHERE, Hampton Institute
Economic Concerns of Black and White Women: Do They Differ?
JANIS BARRY, Fordham University
Women Production Workers: Hazardous Work and Low Pay
Discussants: ELIZABETH SAVOCA, Smith College
MARILYN K. SPENCER, Butler University

2:30 P.M. STRUCTURAL CHANGES AND TRANSITIONS IN THE INTERNATIONAL ECONOMY OF THE 1920'S (Joint Session with the Economic History Association)
Presiding: LARRY NEAL, University of Illinois-Urbana
Papers: BARRY EICHENGREEN, Harvard University, AND FRANCESCO GIAVAZZI, University of Venice
Inflation, Consolidation or Capital Levy? European Debt Management in the 1920's
ELMUS WICKER, Indiana University
Terminating Hyper-Inflation in the Dismembered Hapsburg Monarchy
N. F. R. CRAFTS, Oxford University, AND MARK THOMAS, University of Virginia
International Trade and Structural Unemployment in Interwar Britain
Discussants: JOHN JAMES, University of Virginia
STEVEN WEBB, University of Michigan

4:30 P.M. THE AMERICAN ECONOMY: PROSPECTS AND POLICIES (Joint Session with the National Association of Business Economists)
Presiding: EDGAR R. FIEDLER, Conference Board
Panel: PHILLIP CAGAN, Columbia University
ROBERT G. DEDERICK, Northern Trust Company
SAUL H. HYMANS, University of Michigan

8:00 P.M. RICHARD T. ELY LECTURE
Presiding: CHARLES P. KINDLEBERGER, Massachusetts Institute of Technology and Brandeis University
Paper: SIR ALEC CAIRNCROSS, St. Peters College, Oxford
Economics in Theory and Practice

Saturday, December 29, 1984

8:00 A.M. CURRENT ISSUES IN INFLATION
Presiding: MAHMOOD A. ZAIDI, University of Minnesota
Papers: DAVID MCCLAIN, Boston University
The Inflation Deceleration: A Meltdown of the Core, or Just Luck?
P. S. ANDERSEN, BIS, Switzerland
Real Wages, Inflation and Unemployment
ALEX CUKIERMAN, Carnegie-Mellon University, AND ZVI HERCOWITZ, Tel-Aviv University
Inflation, the Structure of Financial Intermediation and the Real Interest Rate
Discussants: CALVIN D. SIEBERT, University of Iowa
MAHMOOD A. ZAIDI, University of Minnesota

8:00 A.M. LAW AND LABOR UNIONISM IN THE PUBLIC SECTOR (Joint Session with Public Choice Society)
Presiding: ALBERT E. REES, Sloan Foundation
Papers: JOSEPH D. REID, George Mason University, AND MICHAEL M. KURTH, McNeese State University
The Logic of Labor Legislation in the Public Sector
RICHARD FREEMAN AND BERNARD ICHNIOW, Harvard University
Effect of Public Sector Collective Bargaining Laws on Unionization
ROBERT G. EHRENBERG AND ROBERT S. SMITH, Cornell University
Comparable Worth in the Public Sector
Discussants: GORDON TULLOCK, *Public Choice*
JAMES BENNETT, *The Journal of Labor Research*

8:00 A.M. EQUITY IMPLICATIONS OF PUBLIC POLICY IN AN UNSTABLE AGRICULTURAL ECONOMY (Joint Session with the American Agricultural Economics Association)
Presiding: VERNON RUTTAN, University of Minnesota
Papers: GERSHON FEDER, World Bank
The Role of Policy in Diffusion of Technology with Heterogeneous Agriculture

WALLACE HUFFMAN, Iowa State University
Human Capital Adaptive Ability and the Distributional Implications of Agricultural Policy
RICHARD E. JUST AND DAVID ZILBERMAN, University of California-Berkeley
Risk Aversion, Technology Choice, and Equity Effects of Agricultural Policy
Discussants: ROBERT EVENSON, Yale University
WILLIS PETERSON, University of Minnesota

8:00 A.M. BECKER ON CRIME, CHILDREN, AND LOVE
Presiding: SAMUEL L. MYERS, University of Pittsburgh
Papers: CHARLOTTE D. PHELPS, Temple University
Caring and Family Income
ANDREW J. BUCK, SIMON HAKIN, Temple University, AND Z. WEINBLATT, Ben-Gurion University
Moral Sensitivity of Individual Criminals
LEE E. EDLEFSEN, University of Washington
Children as Investments: Some Economic Implications for the Economic Theory of Fertility
Discussants: GARY S. BECKER, University of Chicago
MICHAEL S. PIORE, Massachusetts Institute of Technology

8:00 A.M. THE USE AND ABUSE OF ECONOMETRICS
Presiding: EDWARD E. LEAMER, University of California-Los Angeles
Papers: ZVI GRILICHES, Harvard University
Data Problems in Econometrics
DONALD MCCLOSKEY, University of Iowa
The Loss Function Has Been Mislaid: The Rhetoric of Tests of Significance in Econometrics
JOHN GEWEKE, Duke University
Macroeconometric Modelling and the Theory of the Representative Agent
Discussants: DALE W. JORGENSON, Harvard University
A. H. STUDENMUND, Occidental College

8:00 A.M. TECHNOLOGICAL CHANGE, TRANSFERS, AND REGIONAL PRODUCTIVITY
Presiding: CHARLES R. HULTEN, The Urban Institute
Papers: ROLAND L. MOOMAW, Oklahoma State University, AND MARTIN WILLIAMS, Northern Illinois University
Factor-Augmenting Technical Progress and Productivity in U.S. Manufacturing, 1954 to 1976—A Regional Analysis
SCOTT A. MONROE, Oakland University
Geographic Productivity Experiences of the U.S. Manufacturing Industries from 1964 to 1976
GAUTAM BHATTACHARYYA, University of Kansas
Technology Transfer, Licensing and the Speed of Endogenous Technological Change
Discussants: (To be announced)

8:00 A.M. MODELING INTER-COUNTRY LINKAGES
Presiding: KARL KRESL, Bucknell University
Papers: JACOB COHEN AND STEVEN HUSTED, University of Pittsburgh
An Integrated Accounting Matrix for Canada and the United States
CLARK W. REYNOLDS, Stanford University
Modeling U.S.-Mexican Linkages
LAWRENCE R. KLEIN AND BENJAMIN FRANKLIN, University of Pennsylvania
The Status of Project Link
Discussants: GUILLERMO ORTIZ, Bank of Mexico
JOHN A. SAWYER, University of Toronto

8:00 A.M. DETERMINANTS OF HEALTH STATUS (Joint Session with the Health Economics Research Organization)
Presiding: BARBARA H. KEHRER, Robert Wood Johnson Foundation
Papers: DOUGLAS ELFNER AND K. CELESTE GASPARI, University of Vermont
Estimating the Impact of an Unobservable Variable: The Impact of Lifestyle on Health Status
HARRIET DULEEP, Social Security Administration
Work and Longevity
J. P. NEWHOUSE, H. BAILIT, R. H. BROOK, R. VALDEZ, AND J. E. WARE, Rand Corporation
Health Status Results from the Health Insurance Experiment
Discussants: MICHAEL GROSSMAN, National Bureau of Economic Research
RICHARD BURKHAUSER, Vanderbilt University
VICTOR FUCHS, Stanford University

8:00 A.M. EMPLOYMENT AND WAGE DIFFERENTIAL
Presiding: SHERWIN ROSEN, University of Chicago
Papers: ANDREA BELLER AND FRANCINE BLAU, University of Illinois-Urbana
Trends in the Male-Female Earnings Differential: 1971–81
JAMES B. KURISH, University of Hartford
Compensating Wage Differentials: An Empirical Test of their Existence and Impact on the Union/Nonunion Wage Differential
CLIVE BULL AND MARY MCCARTHY, New York University
Layoffs, Labor Hoarding, and Search
Discussants: DANIEL HAMERMESH, Michigan State University
SHELLY J. LUNDBERG, University of Pennsylvania

8:00 A.M. MERGERS IN ECONOMIC THEORY AND PRACTICE
Presiding: KENNETH ELZINGA, University of Virginia
Papers: PABLO T. SPILLER, University of Pennsylvania
On Vertical Mergers
ALAN A. FISHER AND ROBERT H. LANDE, Federal Trade Commission
Ascertaining Efficiency and Price Effects of Mergers in a Litigation Context: Is it Possible?
GEORGE BITTLINGMAYER, Federal Trade Commission and University of Michigan
Mergers, the Organization of Economic Activity, and Inflation
Discussants: ROBERT STILLMAN, Lexecon, Inc.
KENNETH ELZINGA, University of Virginia
MICHAEL GORT, State University of New York-Buffalo

8:00 A.M. THE UNITED STATES AND THE ASIAN COUNTRIES: ECONOMIC INTERDEPENDENCE (Joint Session with the Committee on Asian Economic Studies)
Presiding: MANORANJAN DUTTA, Rutgers University
Papers: YIEN-I TU, University of Arkansas
Interdependence Between the Developed and Developing Countries: The Case of the United States, Taiwan, and Japan
SALEEM M. KHAN, Bloomsburg University of Pennsylvania
A Case Study of Trade Flows Between the United States and South Asian Countries: Pakistan, India, and Bangladesh
BONG JOON YOON, State University of New York-Binghamton
Production Inefficiences in Two East Asian Economies: South Korea and Taiwan—Its Implications
Discussants: JAMES T. H. TSAO, U.S. International Trade Commission
GEORGE ROSEN, University of Illinois-Chicago
RICHARD HOOLEY, University of Pittsburgh

8:00 A.M. PEACE ECONOMICS II: POLITICAL ECONOMY OF NATIONAL SECURITY AND DEMILITARIZATION (Joint Session with the Peace Science Society (International))
Presiding: JULIUS MARGOLIS, University of California-Irvine
Papers: RICHARD NELSON, Yale University
Political Economy of the National Security System
JANICE MADDEN, University of Pennsylvania
The Effect of Demilitarization on the U.S. Labor Market

10:15 A.M. TOPICS IN THE THEORY OF INDUSTRIAL ORGANIZATION
Presiding: HOWARD KUNREUTHER, University of Pennsylvania
Papers: JOSEPH E. HARRINGTON, JR., Duke University
Entry Deterrence When the Incumbent Firm Has Better Information on Cost
MO-YIN TAM, University of Illinois-Chicago Circle
A Mechanism to Induce Ramsey Pricing for Natural Monopoly Firms
Discussants: MARTIN GAYNOR, University of Texas-Arlington
MARIUS SCHWARTZ, U.S. Department of Justice

10:15 A.M. REAL INTEREST RATES
Presiding: LELAND YEAGER, University of Virginia
Papers: WILLIAM T. CAVIN, Federal Reserve Bank of Cleveland, AND NICHOLAS KARAMOUZIS, Case Western Reserve University
Monetary Policy and Real Interest Rates: New Evidence from the Money Supply Announcements
DEMETROIS S. GIANNAROS AND BHARAT R. KOLLURI, University of Hartford
The Impact of Federal Deficits and Debt on Interest Rates: An International Empirical Investigation

THORVALDUR GYLFASON, University of Stockholm
Consumption, Saving, and Real Interest: Empirical Survey and Synthesis
Discussants: ALAN A. RABIN, University of Tennessee-Chattanooga
JOHN H. WOOD, Northwestern University

10:15 A.M. IN HONOR OF STEPHEN H. HYMER: THE FIRST QUARTER CENTURY OF THE THEORY OF FOREIGN DIRECT INVESTMENT
Presiding: ALAN M. RUGMAN, Dalhousie University
Papers: JOHN M. DUNNING, University of Reading, AND ALAN M. RUGMAN, Dalhousie University
The Influence of Hymer's Dissertation on the Theory of Foreign Direct Investment
DAVID J. TEECE, University of California-Berkeley
The Contribution of Industrial Organization to the Theory of the International Firm
DONALD J. LECRAW, University of Western Ontario
Hymer's Influence on Public Policy in LDCs
Discussants: RAYMOND VERNON, Harvard University
ROBERT Z. ALIBER, University of Chicago
PAUL STREETEN, Boston University

10:15 A.M. PAYMENTS AND MACROECONOMIC PROBLEMS IN LATIN AMERICA
Presiding: CARLOS J. DIAZ-ALEJANDRO, Columbia University
Papers: GUILLERMO CALVO, Columbia University
Liberalization Experiences in Argentina
SEBASTIAN EDWARDS, University of California-Los Angeles
Liberalization with Stabilization in Chile: 1973–83
LANCE TAYLOR, Massachusetts Institute of Technology
Macroeconomic Policy Problems in Mexico
Discussants: JAMES HANSON, World Bank
EDWARD BUFFIE, University of Pennsylvania

10:15 A.M. ECONOMIC FLOWS TO CHILDREN
Presiding: BURT BARNOW, U.S. Department of Labor
Papers: BARBARA BERGMANN, University of Maryland
Structures Insuring Economic Flows to Children
NANCY FOLBRE, New School for Social Research
Rotten Kids and Sugar Daddies: Intergenerational Conflict and Fertility Decline
NANCY S. BARRETT, American University
The Welfare Trap
Discussants: W. LEE HANSEN, University of Wisconsin-Madison
DANIEL SAKS, Vanderbilt University

10:15 A.M. COST BENEFIT AND COST-EFFECTIVENESS STUDIES OF HEALTH CARE PROGRAMS (Joint Session with the Health Economics Research Organization)
Presiding: DONALD E. YETT, University of Southern California
Papers: JON B. CHRISTIANSON AND RONALD J. VOGEL, University of Arizona
Evaluating Investments in Health Care Programs in Unstable Political Environments: The Case of El Salvador
ARNOLD H. RAPHAELSON AND CHARLES P. HALL, JR., Temple University
A Model of Relative Cost Effectiveness of Intermediate Alternatives: Outpatient vs. Inpatient Alcohol Detoxification
BURTON A. WEISBROD, University of Wisconsin-Madison, AND JOHN H. HUSTON, Trinity University
Assessing the Benefits and Costs of Human Vaccines
Discussants: RALPH L. ANDREANO, University of Wisconsin-Madison
MICHAEL D. INTRILIGATOR, University of California-Los Angeles
JOHN M. MARSHALL, University of California-Santa Barbara

10:15 A.M. DEMAND FOR ELECTRICITY AND OTHER ENERGY
Presiding: DANIEL MCFADDEN, Massachusetts Institute of Technology
Papers: D. W. CAVES, L. R. CHRISTENSEN, AND J. A. HERRIGES, Wisconsin Economic Research Institute
Residential Demand for Electricity: Evidence from a Controlled Experiment
CHRISTOPHER GARBACZ, University of Missouri-Rolla
Residential Electricity Demand: Evaluating Functional Form and Structure

PATRICE C. IGNELZI, Cambridge Systematics, Inc.
The Impact of Energy Conservation Programs of Electricity and Gas Companies
Discussants: PAUL L. JOSKOW, Massachusetts Institute of Technology
DANIEL MCFADDEN, Massachusetts Institute of Technology

10:15 A.M. MONOPOLY AND INTERNATIONAL TRADE
Presiding: JAGDISH BHAGWATI, Columbia University
Papers: GEORGE SWEENEY AND WOO JUNG, Vanderbilt University
Import Restrictions and Oligopoly Interactions
R. GLENN HUBBARD, Northwestern University, AND ROBERT J. WEINER, Harvard University
International Stockpiling Agreements as Responses to Commodity Shocks
BRUCE LARSON, University of North Carolina-Asheville
Bickerdike's Contributions to Economics
Discussants: BARBARA SPENCER, Boston College
WOLFGANG MAYER, University of Cincinnati

10:15 A.M. THEORETICAL ANACHRONISM OR CENTRAL CONCEPT FOR ENVIRONMENTAL POLICYMAKING (Joint Session with the Association of Environmental and Resource Economists)
Presiding: V. KERRY SMITH, Vanderbilt University
Papers: ANTHONY C. FISHER AND W. MICHAEL HANEMANN, University of California-Berkeley
Option Value and the Problem of Endangered Species
A. MYRICK FREEMAN III, Bowdoin College
Uncertainty and Environmental Policy: The Role of Option and Quasi-Option Values
RICHARD C. BISHOP, University of Wisconsin-Madison
The Theoretical and Empirical Relevance of Option Value for Environmental Policy
Discussants: WILLIAM D. SCHULZE, University of Colorado
ROBERT MENDELSOHN, University of Washington and University of Michigan
V. KERRY SMITH, Vanderbilt University

10:15 A.M. ECONOMIC HISTORY: A NECESSARY THOUGH NOT SUFFICIENT CONDITION FOR AN ECONOMIST (Joint Session with the Economic History Association)
Presiding: W. ARTHUR LEWIS, Princeton University
Panel: KENNETH J. ARROW, Stanford University
PETER TEMIN, Massachusetts Institute of Technology
ROBERT M. SOLOW, Massachusetts Institute of Technology
PAUL A. DAVID, Stanford University
Discussants: ALBERT FISHLOW, University of California-Berkeley
DONALD N. MCCLOSKEY, University of Iowa
GAVIN WRIGHT, Stanford University

10:15 A.M. FEDERAL RESERVE, CENTRAL BANKING, AND ECONOMIC ACTIVITY (Joint Session with the Eastern Economic Association)
Presiding: ANDREW F. BRIMMER, Brimmer & Co., Inc.
Papers: ALLEN SINAI, Shearson Lehman/American Express
The New Federal Reserve Policy: An Appraisal
HYMAN MINSKY, Washington University-St. Louis
Central Banking and Money Market Changes: A Reprise
ANDREW F. BRIMMER, Brimmer & Co., Inc.
The Federal Reserve as a Lender of Last Resort
Discussants: RONALD G. BODKIN, University of Ottawa
DUDLEY DILLARD, University of Maryland
IRA O. SCOTT, New York Institute of Technology

10:15 A.M. ECONOMIC AND BUSINESS OUTLOOK (Joint Session with the National Association of Business Economists)
Presiding: BEN E. LADEN, T. Rowe Price Associates, Inc.
Papers: F. THOMAS JUSTER, University of Michigan
Outlook for Consumer Spending
RICHARD D. RIPPE, Dean Witter Reynolds, Inc.
Prospects for Capital Spending
DAVID HALE, Kemper Financial Services, Inc.
International Risks to the Outlook
Discussants: JAMES F. SMITH, Union Carbide Corporation
DONALD H. STRASZHEIM, Wharton EFA, Inc.

10:15 A.M. DISCRIMINATION AND INCOME INEQUALITY: REPARATIONS, REDISTRIBUTION AND RACE: IS THERE A SOCIAL DEBT? (Joint Session with the National Economic Association)
Presiding: RICHARD F. AMERICA, U.S. Small Business Administration
Papers: SHELDON DANZIGER, University of Wisconsin-Madison
(To be announced)
LESTER THUROW, Massachusetts Institute of Technology
(To be announced)
STANLEY MASTERS, State University of New York-Binghamton
(To be announced)
DAVID SWINTON, Clark College
(To be announced)
Discussants: PETER LINDERT, University of California-Davis
DONALD HARRIS, Stanford University

10:15 A.M. DEVELOPMENT STRATEGIES FOR ACCELERATED WORLD DEVELOPMENT (Joint Session with the Society for Policy Modeling)
Presiding: DOMINICK SALVATORE, Fordham University
Papers: STANLEY J. LAWSON, St. John's University, AND DOUGLAS O. WALKER, United Nations
Past Trade Patterns and the Future Potential of Increased Intra-Developing Country Trade
IRMA ADELMAN, University of California-Berkeley
Development Strategies for Equitable World Development
GRACIELA CHICHILNISKY, Columbia University
Resources in the Global Economy
Discussants: STEVEN K. CHANG, Kennesaw College
EVANGELOS S. DJIMPOULOS, Fairleigh Dickinson University
GEORGE C. WANG, California State University

12:30 P.M. LUNCHEON IN HONOR OF NOBEL LAUREATE GERARD DEBREU
Presiding: CHARLES L. SCHULTZE, The Brookings Institution
Speaker: KENNETH J. ARROW, Stanford University

2:30 P.M. ECONOMIC EDUCATION: THE USE OF COMPUTERS
Presiding: W. LEE HANSEN, University of Wisconsin-Madison
Papers: JOHN SUMANSKY, Joint Council on Economic Education
Computer Applications in Pre-College Economics
RAY C. FAIR, Yale University, AND KARL E. CASE, Wellesley College
Macro Simulations for PCs in the Classroom
DARRELL R. LEWIS AND BRUCE R. DALGAARD, University of Minnesota
Cost Effectiveness of Computer Assisted Economics Instruction
Discussants: JAMES L. POWELL, Massachusetts Institute of Technology
SUSAN K. FEIGENBAUM, Claremont McKenna College and Claremont Graduate School

2:30 P.M. THE END OF THE GREAT BOOM AND THE BREAKDOWN OF BRETTON WOODS: WAS IT A COINCIDENCE?
Presiding: CHARLES P. KINDLEBERGER, Massachusetts Institute of Technology and Brandeis University
Papers: JOHN BILSON, University of Chicago
Microeconomic Instability and Flexible Exchange Rates
J. CARTER MURPHY, Southern Methodist University
Optimal Constraints on National Policies
JOHN WILLIAMSON, Institute for International Economics
On the System in Bretton Woods
Discussants: MATTHEW CANZONERI, Federal Reserve Board
PAUL KRUGMAN, Massachusetts Institute of Technology

2:30 P.M. MONETARY PAPERS AS THE WORLD CHANGES
Presiding: RICHARD STARTZ, University of Washington
Papers: JESSE ABRAHAMS AND DAVID WYSS, Data Resources, Inc.
Macro Effects of the Housing Market Deregulation
JOHN DOUKAS, Concordia University
The Rationality of Money Supply Expectations and Exchange Rate Responses to Money Supply Announcements
ELINOR HARRIS SOLOMON, George Washington University
Monetary vs. Interest Rate Targets Revisited: The Role of Market Expectations
Discussants: STEPHEN MAYER, Federal Reserve Bank of Philadelphia
JOHN HUIZINGA, University of Chicago

2:30 P.M. THE GOVERNMENT AND SAVING
Presiding: WILLIAM J. BEEMAN, Congressional Budget Office
Papers: EDWARD MONTGOMERY, Federal Reserve Board
Observations on the Recent Decline in the Personal Saving Rate
CHARLES STEINDEL, First National Bank of Chicago
The Government and Saving: Basic Issues
DARREL COHEN, Federal Reserve Board, AND PETER SKAPERDAS, Federal Reserve Bank of New York
The Size of the Public Sector: Implications for Growth and Performance
Discussants: FRANK DE LEEUW, U.S. Bureau of Economic Analysis
HARVEY S. ROSEN, Princeton University

2:30 P.M. EMPIRICAL MODELS OF LONG-TERM CONTRACTS: ADVANCES IN LABOR ECONOMICS
Presiding: GEORGE NEUMANN, Northwestern University and University of Iowa
Papers: LAURENCE J. KOTLIKOFF, Boston University, AND DAVID WISE, Harvard University
Labor Compensation and the Structure of Private Pension Plans: Evidence for Contractual vs. Spot Markets
OLIVIA S. MITCHELL AND REBECCA LUZADIS, Cornell University
Firm-Level Policy Toward Older Workers
JOHN ABOWD, University of Chicago, AND DAVID CARD, Princeton University
Intertemporal Substitution in the Presence of Long-Term Contracts
Discussants: STEVEN G. ALLEN, North Carolina State University
GEORGE NEUMANN, Northwestern University and University of Iowa

2:30 P.M. THE AFTER-KEYNES CAMBRIDGE CONTRIBUTIONS TO THEORY
Presiding: HYMAN P. MINSKY, Washington University-St. Louis
Papers: JAN KREGEL, Groningen University, The Netherlands
Hamlet Without the Prince: Cambridge Macroeconomics Without Money
BERTRAM SCHEFOLD, Johann Wolfgang Goethe University
Cambridge Price Theory: Special Model or General Theory of Value
E. ROY WEINTRAUB, Duke University
Critiques of Equilibrium: An Appraisal
Discussants: DONALD HARRIS, Stanford University
MARK KUPENBERG, Swarthmore College

2:30 P.M. THE CONSEQUENCES OF DEREGULATION IN THE TRANSPORTATION AND TELECOMMUNICATIONS SECTORS
Presiding: PAUL L. JOSKOW, Massachusetts Institute of Technology
Papers: JOHN MEYER, Harvard University, AND WILLIAM B. TYE, Putnam, Hayes & Barlett, Inc.
The Regulatory Transition
ROGER NOLL, California Institute of Technology
State Responses to Federal Telecommunications Deregulation
STEVEN MORRISON, Northeastern University, AND CLIFFORD WINSTON, Massachusetts Institute of Technology and The Brookings Institution
The Welfare Effects of Airline Deregulation
Discussants: ROBERT WILLIG, Princeton University
THOMAS MOORE, Hoover Institution

2:30 P.M. TOPICS IN FISCAL FEDERALISM
Presiding: PETER MIESZKOWSKI, Rice University
Papers: CHARLES C. BROWN AND WALLACE E. OATES, University of Maryland
Assistance to the Poor in a Federal System
K. HAYES, Northern Illinois University, S. GROSSKOPF, Southern Illinois University, AND S. PORTER-HUDAK, Northern Illinois University
The Effect of Pension and Funding Practices on Municipal Labor Costs
ROBERT P. INMAN, University of Pennsylvania and National Bureau of Economic Research
Local Taxes and Local Debt: The Political Economy of the Fiscal Crisis
Discussants: EDWARD GRAMLICH, University of Michigan
GEORGE PETERSON, The Urban Institute

2:30 P.M. INNOVATION, PRODUCTIVITY, AND INDUSTRIAL COMPETITIVENESS
Presiding: ROMESH DIWAN, Rensselaer Polytechnic Institute
Papers: LEO SVEIKAUSKAS, Rensselaer Polytechnic Institute
Major Industrial Innovations: Research and Development and Productivity Growth

MARTIN N. BAILY, The Brookings Institution, AND ALOK CHAKRABARTI, Drexel University
Innovation, Productivity, and Industrial Competitiveness
GERHARD O. MENSCH, Case Western Reserve University
The Interaction of Product and Process Innovation with Short-Term Fluctuations in Industry
Discussants: F. M. SCHERER, Swarthmore College
ROLF PIEKARZ, National Science Foundation

2:30 P.M. WORK ORGANIZATION AND CONTROL TECHNOLOGY
Presiding: PETER S. ALBIN, John Jay College, City University of New York
Papers: DAVID LANDES, Harvard University
Work Organization Practices in Manufacturing Industry
PETER S. ALBIN, John Jay College, City University of New York
Job Design, Control Technology and Technical Change
MICHAEL S. PIORE, Massachusetts Institute of Technology
New Systems of Work Organization
Discussant: PETER UNDERWEGER, United Auto Workers

2:30 P.M. ASIAN ECONOMIC DEVELOPMENT (Joint Session with the Committee on Asian Economic Studies)
Presiding: LAWRENCE R. KLEIN, University of Pennsylvania
Papers: HIROSHI KITAMURA, International University of Japan
Japan and Asia-Pacific Economic Integration
PAUL P. STREETEN, Boston University
New Directions for Private Resource Transfers
E. WAYNE NAFZIGER, Kansas State University
The Japanese Development Model: Its Implications for Less Developed Countries
Discussants: GARY SAXONHOUSE, University of Michigan
SHANTI S. TANGRI, Rutgers University
JOHN M. LETICHE, University of California-Berkeley

2:30 P.M. THE OUTLOOK FOR FINANCIAL MARKETS (Joint Session with the National Association of Business Economists)
Presiding: A. NICHOLAS FILIPPELLO, Monsanto Company
Papers: W. LEE HOSKINS, Pittsburgh National Bank
Federal Reserve Policy and the Outlook for Interest Rates
DAVID M. JONES, Aubrey G. Lanston & Company
Treasury Financing and the Impact on the Flow of Funds
BEN E. LADEN, T. Rowe Price Associates, Inc.
The Environment for the Stock Market
Discussants: KATHLEEN M. COOPER, Security Pacific National Bank
RICHARD D. RIPPE, Dean Witter Reynolds, Inc.

4:30 P.M. PRESIDENTIAL ADDRESS
Presiding: ALICE M. RIVLIN, The Brookings Institution
Speaker: CHARLES L. SCHULTZE, The Brookings Institution

5:30 P.M. BUSINESS MEETING

Sunday, December 30, 1984

8:00 A.M. YESTERYEARS' LONG-RANGE PROJECTIONS: A RETROSPECTIVE
Presiding: RAYMOND VERNON, Harvard University
Papers: DICK NETZER, New York University
1985 Projections of the New York Metropolitan Region Study
WILLIAM R. CLINE, Institute for International Economics
Selected Projections Made in the 1950's with Respect to Future International Economic Developments
JENNIFER ROBACK, Yale University
On the Implicit Economic Projections of Orwell for 1984
Discussants: BENJAMIN CHINITZ, University of Lowell
WALTER S. SALANT, The Brookings Institution

8:00 A.M. IS GENDER EQUALITY ADVANCING IN THE WORK PLACE?
Presiding: ELYCE J. ROTELLA, Indiana University

Papers: HELEN YOUNGELSON, Portland State University
Gender Differences in Salary and Career Advancement: The Experience of Officers in a Single Bank
ROBIN L. BARTLETT AND TIMOTHY I. MILLER, Denison University
Executive Compensation: The Female Executive Risk Taking and Clubs
SHARON BERNSTEIN MEGDAL AND MICHAEL R. RANSOM, University of Arizona
Longitudinal Changes in Salary at a University: The Response to Equal Pay Legislation
MARY E. CORCORAN AND PAUL N. COURANT, University of Michigan
A Model of the Labor Market Consequences of Sex-Role Socialization
Discussants: SHULAMIT KAHN, University of California-Irvine
LUVONIA J. CASPERSON, Louisiana State University-Shreveport

8:00 A.M. THE POLITICAL ECONOMY OF TARIFFS
Presiding: ROBERT BALDWIN, University of Wisconsin-Madison
Papers: JAGDISH BHAGWATI, Columbia University
Tariffs and Welfare
STEPHEN P. MAGEE, University of Texas-Austin, WILLIAM A. BROCK, University of Wisconsin-Madison, AND LESLIE YOUNG, University of Texas-Austin
Endogenous Tariff Theory
JOHN A. C. CONYBEARE, Columbia University
National Tariff Levels: Theory and Evidence
Discussants: LYNN GILLETTE, University of Texas-Austin
CHARLES ROWLEY, George Mason University
EDWARD J. RAY, Ohio State University

8:00 A.M. RISK MANAGEMENT AND ENVIRONMENTAL POLICYMAKING (Joint Session with the Association of Environmental and Resource Economists)
Presiding: V. KERRY SMITH, Vanderbilt University
Papers: ALBERT NICHOLS, U.S. Environmental Protection Agency, AND RICHARD ZECKHAUSER, Harvard University
Risk Analysis for Environmental Decisions
PAUL R. PORTNEY AND JOHN MULLAHY, Resources for the Future
Relative Risk Assessment and Air Quality Regulation
PETER HUBER, U.S. Supreme Court
The Old-New Risk Dichotomy and Environmental Policy
Discussants: ROGER NOLL, California Institute of Technology
MILTON RUSSELL, U.S. Environmental Protection Agency
W. KIP VISCUSI, Duke University

8:00 A.M. MODELING PROSPECTS FOR WORLD GROWTH (Joint Session with the Society for Policy Modeling)
Presiding: STANLEY J. LAWSON, St. John's University
Papers: DOMINICK SALVATORE, Fordham University
The New Protectionist Threat to World Welfare
JAIME DE PINIES AND SHAKROKH FARDOUST, United Nations
Structural Interdependence within a Mini North-South Model: The Short and Long Run
FRED CAMPANI, United Nations
Modeling the Future Pattern of Income Distribution in Africa
Discussants: GRACIELA CHICHILNISKY, Columbia University
DOUGLAS O. WALKER, United Nations
GEORGE C. WANG, California State University

8:00 A.M. UNCERTAINTY, BEHAVIOR, AND ECONOMIC THEORY
Presiding: KENNETH J. ARROW, Stanford University
Papers: RONALD A. HEINER, Institute for Advanced Study and Brigham Young University
Origin of Predictable Behavior: Further Modeling and Applications
PETER NEZ, VERNON L. SMITH, University of Arizona, AND ARLINGTON WILLIAMS, Indiana University
Individual Rationality, Market Rationality, and Value Estimation
L. DWIGHT ISRAELSEN AND ALLEN D. LEBARON, Utah State University
Uncertainty and Behavior in Socioeconomic Institutions
Discussants: AXEL LEIJONHUFVUD, University of California-Los Angeles
MARK J. MACHINA, University of California-San Diego

8:00 A.M. ISSUES IN THE ECONOMICS OF *R&D*
Presiding: RICHARD J. GILBERT, University of California-Berkeley
Papers: TIMOTHY F. BRESNAHAN, Stanford University
New Product Competition in Plain Paper Copiers
RICHARD LEVIN, Yale University
Appropriability, Technological Superiority, and *R&D* Spending: Some New Evidence
CARL SHAPIRO, Princeton University
Patent Licensing and *R&D* Rivalry
Discussants: PAUL DAVID, Stanford University
TRACY LEWIS, University of British Columbia

8:00 A.M. HUMAN CAPITAL AND CULTURE: ANALYSES OF VARIATIONS IN LABOR MARKET PERFORMANCE
Presiding: RHONDA M. WILLIAMS, University of Texas-Austin
Papers: NIGEL TOMES, University of Western Ontario
The Effects of Religion and Denomination on Earnings and the Returns to Human Capital
CORDELIA REIMERS, Hunter College, City University of New York
Cultural Differences in Labor Supply Among Hispanic, Black, and Anglo Married Women
WILLIAM A. DARITY, JR., University of North Carolina-Chapel Hill, AND RHONDA M. WILLIAMS, University of Texas-Austin
Culture, Class, and the Production of Human Capital
Discussants: BARRY CHISWICK, University of Illinois-Chicago
STEPHEN STEINBERG, Queens College, City University of New York

8:00 A.M. ACCOUNTING AND ECONOMICS (Joint Session with the Association of Managerial Economics)
Presiding: MARK HIRSCHEY, University of Colorado-Denver
Papers: ROBERT A. CONNOLLY, University of North Carolina-Greensboro
Accounting Profits, Market Value, and Capital Investment
LINDA ELIZABETH DEANGELO, University of Rochester
Accounting Measurement and Managerial Incentives: A Property Rights Approach
RICHARD T. CASTANIAS, DOUGLAS C. CERF, AND PAUL A. GRIFFIN, University of California-Davis
The Effects of Foreign Currency Translation Accounting on Security Analyst Forecasts
Discussants: NICHOLAS DOPUCH, Washington University-St. Louis
D. PAUL NEWMAN, University of Texas-Austin

8:00 A.M. ISSUES IN ECONOMIC DEVELOPMENT (Joint Session with the Committee on Asian Economic Studies)
Presiding: PAUL STREETEN, Boston University
Papers: HARRIET DULEEP, U.S. Department of Health and Human Services
Serfdom-Slavery and the Labor-Land Ratio: A Partial Explanation of the Historical Development of India's Caste System
PREM S. LAUMAS, Northern Illinois University
Monetization, Economic Development, and the Exogeneity of Money
ANTON D. LOWENBERG, Simon Fraser University
An Economic Theory of the Apartheid State
Discussants: JEFFREY JAMES, Boston University
BRIAN LEVY, Williams College

10:15 A.M. THE DEREGULATION OF BANKING IN THE UNITED STATES
Presiding: THOMAS M. HUMPHREY, Federal Reserve Bank of Richmond
Papers: RICHARD H. TIMBERLAKE, JR., University of Georgia
Deregulation and Federal Reserve Policy
LELAND YEAGER, University of Virginia
A Proposal for a Fully Deregulated Monetary and Financial System
LEONARD W. RAPPING, University of Massachusetts-Amherst, AND LAWRENCE B. PULLEY, Brandeis University
Bank Deregulation and the Interest Rate
Discussant: WILLIAM BERANEK, University of Georgia

10:15 A.M. EFFECTS OF HEALTH INSURANCE ON DEMAND FOR HEALTH SERVICES (Joint Session with the Health Economics Research Organization)
Presiding: DONALD E. YETT, University of Southern California
Papers: CURT D. MUELLER AND ALAN C. MONHEIT, National Center for Health Services Research
Insurance Coverage and the Demand for Dental Care

J. MATTHIAS GRAF VON DER SCHULENBERG, International Institute of Management, Berlin
Deregulation of Statutory Health Insurance Protection: The Effects of Increased Cost-Sharing Arrangements
PAMELA J. FARLEY, National Center for Health Services Research
Hospital and Ambulatory Services Used in the Treatment of Selected Illnesses
Discussants: WILLARD G. MANNING, Rand Corporation
MARK V. PAULY, University of Pennsylvania
ROGER FELDMAN, University of Minnesota

10:15 A.M. NEW APPROACHES TO ANALYSIS OF POVERTY IN LDCs
Presiding: GIAN S. SAHOTA, Vanderbilt University and United Nations
Papers: OSCAR ALTIMIR, ECLA, Santiago, Chile
Poverty in Latin America: An Overview
GIAN S. SAHOTA, Vanderbilt University and United Nations, AND JUAN LUIS MORENO, Ministry of Planning, Panama
An Analysis of Panama's Policies from the Viewpoint of Poverty
ROBERT EVENSON, Yale University
Fertility, Child Health, and Poverty in Panama
Discussants: T. N. SRINIVASAN, Yale University and World Bank
RICHARD BURKHAUSER, Vanderbilt University

10:15 A.M. INTERNATIONAL TRADE THEORY
Presiding: RONALD W. JONES, University of Rochester
Papers: FRANCISCO RIVERA-BATIZ, University of Massachusetts-Amherst
Economies of Scale, Increased Competition, and the Gains from Intraindustry Trade
RICK JENSEN AND MARIE THURSBY, Ohio State University
A Decision Theoretic Approach to Technology Transfer
RICHARD POMFRET, Johns Hopkins University
Discrimination in International Trade
SHABTAI DONNENFELD, University of Haifa and Indiana University
Imperfect Information and Trade in Differentiated Products
Discussants: MAKOTO YANO, Cornell University
JOHN POMERY, Purdue University
PAUL JOHNSON, York University

10:15 A.M. INDUSTRIAL POLICY IN FRANCE
Presiding: SUZANNE BERGER, Massachusetts Institute of Technology
Papers: CAGLAR KEYDER, State University of New York-Binghamton
Industrial Policy from Colbert to the Front Populaire
CHRISTIAN STOFFAES, Institut d'Etudies Politiques, Paris
Industrial Policy from the Liberation to Raymond Barre
BELA BALASSA, Johns Hopkins University
Industrial Policy Under the Socialist Government
Discussants: W. JAMES ADAMS, University of Michigan
JOHN ZYSMAN, University of California-Berkeley

10:15 A.M. CURRENT ISSUES IN FOREIGN EXCHANGE
Presiding: JOHN BILSON, University of Chicago
Papers: MICHAEL MELVIN AND DON SCHLAGENHAUF, Arizona State University
Market Expectations and Country Risk: A New Approach
MOON LEE, University of Saskatoon
Real Interest Rates, Taxes, and Flexible Exchange Rates: Canada and the United States
MICHAEL MELVIN, Arizona State University
The Choice of an Exchange Rate System and Macroeconomic Stability: Theory and Evidence

10:15 A.M. MONETARY HISTORY: NEW PERSPECTIVES ON TRADITIONAL THEMES (Joint Session with the Economic History Association)
Presiding: JONATHAN HUGHES, Northwestern University
Papers: ANNA J. SCHWARTZ, National Bureau of Economic Research
The Service Flows of Money: A Historical Perspective
MICHAEL D. BORDO, University of South Carolina and National Bureau of Economic Research
Explorations in Monetary History: A Survey of the Recent Literature

GEOFFREY WOOD AND FORREST CAPIE, City University, London
Gold, Empire and Crisis: Some Explorations in U.K. Monetary History
Discussants: PHILLIP CAGAN, Columbia University
BARRY EICHENGREEN, Harvard University

10:15 A.M. CONSUMPTION ECONOMICS
Presiding: ROBERT A. POLLAK, University of Pennsylvania
Papers: DONALD COURSEY AND CHARLES F. MASON, University of Wyoming
Investigations Concerning the Dynamics of Consumer Behavior
M. CINAR, Loyola University of Chicago
An Examination of a Utility-Based Approximation of Comparative Household Welfare
BARBARA A. VATTER, Memphis State University
Women's Economic Power, the Household, and the Market: An Historical Theory of Work
Discussants: ELIZABETH PETERS, Ohio State University
MARJORIE B. MCELROY, Duke University

10:15 A.M. ORIGINS AND PREDECESSORS OF MODERN U.S. RADICAL POLITICAL ECONOMY (Joint Session with the History of Economics Society)
Presiding: HOWARD J. SHERMAN, University of California-Riverside
Papers: E. K. HUNT, University of Utah
Marx, Veblen, and Modern Radical Economics
DANIEL R. FUSFELD, University of Michigan
Radical Economic Thought in the American Labor Movement
HERBERT GINTIS, University of Massachusetts-Amherst
The Origins of Radical Economics in the United States
Discussants: WARREN J. SAMUELS, Michigan State University
ROBERT L. HEILBRONER, New School for Social Research
JOHN B. DAVIS, Michigan State University

10:15 A.M. ANTITRUST AND RACE (Joint Session with the National Economic Association)
Presiding: RICHARD F. AMERICA, U.S. Small Business Administration
Papers: DOUGLAS F. GREER, San Jose State University
(To be announced)
WILLIAM G. SHEPHERD, University of Michigan, AND GAVIN CHEN, U.S. Department of Commerce
(To be announced)
LEONARD WEISS, University of Wisconsin-Madison
(To be announced)
Discussants: F. M. SCHERER, Swarthmore College
WILLIAM COMANOR, University of California-Santa Barbara
SAMUEL L. MYERS, University of Pittsburgh

NOTES

The ninety-seventh annual meeting of the American Economic Association will be held in Dallas, Texas, December 28–30, 1984.

The Professional Placement Service will be located at the Convention Center. It will be open from 10:00 A.M. to 5:00 P.M., December 27; 9:00 A.M. to 5 P.M., December 28–29, and 9:00 A.M. to 12:00 noon, December 30.

Members wishing to give papers or make suggestions for the program for the AEA meeting, New York, NY, December 28–30, 1985, are invited to write to Alice M. Rivlin, The Brookings Institution, 1775 Mass. Ave., NW, Washington, D.C. 20036. To be considered for the contributed sessions, abstracts of (noneconometric) papers must be received no later than February 1, 1985.

The Executive Committee of the Association, meeting on December 27, 1983 in San Francisco, voted to establish a Search and Structure Committee for the *Journal of Economic Literature*. It was asked to recommend a replacement for the current Managing Editor, Moses Abramovitz, who has indicated that he does not wish to serve beyond his current term, which expires at the end of 1985. The Committee was also asked to recommend with respect to the possible reintegration in a single place of editorial functions that, since 1981, have been divided between Stanford and Pittsburgh.

President Charles L. Schultz has now appointed such a Committee, with the following members: Gardner Ackley (Chair), James Buchanan, Alan Blinder, Albert Rees, Kerry Smith, Stanley Black, and Allen Kelley.

The Committee urgently invites formal or informal communications from members, proposing themselves, on recommending others, to serve for a three-year renewable term as Managing Editor of the *JEL*, beginning January 1, 1986. The date of transfer of responsibilities to the new Editor can be negotiated. Communications may be sent to the Secretary of the Association; to Gardner Ackley, Department of Economics, University of Michigan, Ann Arbor, MI 48109; or to another member of the Committee. Such communications will be treated as confidential, if so requested. It is hoped that appointment of the new Editor can be officially accomplished at the Executive Committee Meeting of December 1984, or (at the latest) of March 1985.

The Conference of the Society of Economic and Social Sciences, "The Economics of Health and Medical Care," will be held September 16–18, 1985, in Saarbrücken, West Germany. For further information, contact Professor Dr. Gerard Gäfgen, Universität Konstanz, Fakultät für Wirtschaftswissenschaften und Statistik, Postfach 5560, D-7750 Konstanz 1, FRG (Telephone: 07531-88 25 68/66).

The second International Conference on Consumer Behavior and Energy Policy will be held April 10–12, 1985, in Paris. The focus is on the nontechnical barriers to the rational use of energy in the residential sector. Papers will be based on research from the behavioral, socio-cultural, economic, and policy analytic perspectives. For further details, write to Eric Monnier—MEC2: CSTB, 4 avenue du Recteur Poincaré, 75782 Paris 16, France, or George Gaskell, London School of Economics, Houghton Street, Aldwych, London WC2 2AE, England.

Call for Papers: The Southwestern Economics Association will hold its annual meeting, March 20–23, 1985, in Houston, Texas, in conjunction with the annual meeting of the Southwest Social Science Association. Economists and others with an interest in the discipline are invited to submit proposals for papers and panel discussions, or to serve as chairperson or discussant. Contact Kathie S. Gilbert, President-Elect SEA, Department of Economics and Finance, Mississippi State University, P.O. Drawer JE, Mississippi State, MS 39762.

Call for Papers: The Committee on Asian Economic Studies, an interuniversity program, will hold the second Conference on United States-Asia Economic Relations, May 20–22, 1985, in New York City. Those interested in presenting papers should send a 150-word abstract before November 1, 1984. A complete draft of the final paper is due March 31, 1985. For further information, contact Professor M. Dutta, Director, Committee on Asian Economic Studies, Department of Economics, Rutgers University, New Brunswick, NJ 08903.

Call for Papers: The annual meeting of the Eastern Finance Association will be held April 24–27, 1985, in Williamsburg, Virginia. Those interested in presenting papers should submit a two-page, single spaced abstract by November 1, 1984. Individuals interested in chairing sessions, organizing special sessions, or serving as discussants should write indicating their areas of interest. Contact Diana R. Harrington, Vice-President, EFA Program, Darden Graduate School of Business, University of Virginia, Box 6550, Charlottesville, VA 22906.

Call for Papers: The International Conference on Economic Policies and Control Theory will be held in Venice, Italy, January 29–February 1, 1985. Papers should be submitted in triplicate, with four copies of an abstract giving author's name, address, telephone number, and designation of one of the following fields: linear and nonlinear control; adaptive control; dynamic games; stochastic simulations; estimation and control of

economic and econometric models; model specification and policy effectiveness. The absolute deadline is October 31, 1984. Submit to the Committee Chairman: Professor Giovanni Castellani, Universita' degli Studi di Venezia, Ca' Foscari, 30123 Venezia, Italy.

Call for Papers: Population Research and Policy Review, a journal of Elsevier Science Publishers, seeks papers on the comparable worth doctrine. The journal is concerned with the link between empirical research in the social sciences and public policy bearing on the structure and dynamics of populations; articles are generally nontechnical in nature, focusing on the policy ramifications of research findings. Manuscripts should be submitted in duplicate to Larry D. Barnett, Editor-in-Chief, School of Law, Widener University, P.O. Box 7474, Wilmington, DE 19803-0474.

Call for Papers: The Western Social Science Association will hold its twenty-seventh annual meeting at the Americana Hotel in Fort Worth, Texas, April 14–27, 1985. Proposals should be sent by November 15, 1984 to Ralph Hutchinson, Department of Economics, California State Polytechnic University, Pomona, CA 91768, for micro and other areas, or to Charles J. Ellard, Department of Economics and Accounting, Pan American University, Edinburg, TX 78539, for macro and finance areas.

The Europäische Zeitschrift für Politische Ökonomie (European Journal of Political Economy) will provide a forum for work in classical and neoclassical political economy, public choice and collective decision making, and economic history, and will be published in English and German. Papers in French, Italian, Spanish, Dutch, and the Scandinavian languages are accepted for review and eventual translation. Submit manuscripts to Manfred J. Holler, Department of Economics, University of Munich, 8000 Munich 22, FRG.

The Journal of Energy and Natural Resources Law is cosponsored by the Section on Energy and Natural Resources Law of the International Bar Association and the Centre for Petroleum and Mineral Law Studies. It is the policy of the editors to include articles relevant to the issues of energy and natural resources law written from the standpoint of other disciplines such as economics. Submit manuscripts in duplicate to the Editor, *JENRL*, Centre for Petroleum and Mineral Law Studies, University of Dundee, Dundee DD1 4HN, U.K.

The COPSS Visiting Lecturer Program in Statistics is sponsored by the Committee of Presidents of Statistical Societies and supported jointly by the American Statistical Association, the Biometric Society, the Institute of Mathematical Statics, the Society of Actuaries, and the Statistical Society of Canada. Leading statisticians from universities, industry, and government participate as lecturers. The program is available to colleges, high schools, and other interested groups in the continental United States and Canada. A detailed brochure describing the program is available from Jon R. Kettenring, Chair, Visiting Lecturer Program, 29 Pine Grove Avenue, Summit, NJ 07901.

Nominations for candidates for Sloan Research Fellowships are due by September 15. Candidates must be faculty members at a college or university in the United States or Canada, and must be at an early stage of their research careers. For information, write Sloan Research Fellowships, Alfred P. Sloan Foundation, 630 Fifth Avenue, New York, NY 10111.

The Council for International Exchange of Scholars announces the 1985–86 Fulbright Scholar-in-Residence Program. Community and junior colleges, and four-year colleges and universities may submit proposals to invite a scholar from abroad to lecture for an academic year or term in any field of the humanities or social sciences. Additional information and proposal forms may be obtained from CIES, Eleven Dupont Circle, NW, Washington, D.C. 20036. Deadline is November 1, 1984.

The Mental Health Economics Research Program of the National Institute of Mental Health is soliciting research grant applications. The period of support is one to five years. Some areas of research interest are utilization and costs of mental health services, prospective payment for mental health services, financing/costs of mental health technologies, demand/supply of mental health services, work productivity, and econometric studies. Applications are accepted throughout the year for new grants as well as for competing continuation and supplemental grants. Applicants are invited to seek consultation and assistance in the development of concept papers and drafts of proposals. The deadlines for receipt of applications, reviews, and funding dates vary. For further information, contact Mr. Paul Widem, Assistant Chief, Mental Health Economics Research Branch, Division of Biometry and Epidemiology, National Institute of Mental Health, Parklawn Bldg., Room 18C26, 5600 Fishers Lane, Rockville, MD 20857. (Telephone 310+443–3683)

The Graduate School of Industrial Administration, Carnegie-Mellon University, postdoctoral fellowship provides a unique opportunity for a researcher with a strong commitment to the use of mathematical or quantitative analysis in the study of politics and the interdependence of political and economic decision

making. Fellows are to devote a twelve-month period of residence to research. There are no teaching duties. Fellows may also take advantage of GSIA's doctoral program to obtain additional training in advance topics. Applications are invited from recent Ph.D.s in economics, political science, or related fields. A twelve-month stipend, fully commensurate with an assistant professor's salary, will be provided. Applicants may either forward a resume and brief statement of research interests, or write for further information to Professor Howard Rosenthal, GSIA Carnegie-Mellon University, Pittsburgh, PA 15213.

The American Council of Learned Societies announces 1985–86 fellowship competitions for support for postdoctoral research and for doctoral dissertation research on East Europe (except projects to be undertaken in Eastern Europe) and China (except projects to be undertaken in PRC). Proposals may be in any discipline of the humanities or social sciences; they may be multidisciplinary, and they may be comparative. In requesting application forms, give the following: highest academic degree held and date received; citizenship and permanent residence; academic or other position; subject of proposed research; period for which support will be requested; and specific competition in which application is contemplated. The deadline is November 15, 1984. Further information is available on request from Office of Fellowship and Grants, ACLS, 228 East 45 Street, New York, NY 10017.

The Social Science Research Council announces the 1984 deadlines for 1985–86 research fellowships and grants:

International Doctoral Research Fellowships: applicants must be graduate students in the social sciences, the humanities, or professional fields who will have completed all requirements for the Ph.D. except the dissertation at the time the fellowship is to begin. These fellowships are for research to be carried out in Africa, Asia, Latin America, and the Caribbean, the Near and Middle East, Western Europe, or for cross-area research. Deadline is November 1, 1984.

International Postdoctoral Research Grants: for research in or on Africa, Asia, Latin American and the Caribbean, and the Near and Middle East. They may be used to support research on one country, comparative research between countries within an area, or comparative research between areas. There is also a special program for collaborative research between American and foreign scholars on Latin America. Deadline is December 1, 1984.

Indochina Studies Grants support research, writing, and the archiving of materials on Cambodia, Laos, and Vietnam, drawing on the knowledge and experience of refugees who have left since 1975, and are now residing in North America. Specifically excluded are projects concerned with American experience in Indochina, and the experience of Indochina refugees in North America. Deadline is December 1, 1984.

Write to the Council, 605 Third Avenue, New York, NY 10158. Social Science Research Applications must be submitted on forms provided by the Council.

The Canadian Studies Faculty Enrichment Programme provides selected faculty of four-year U.S. universities and colleges with an opportunity to undertake study relating to Canada in order to develop new courses, or redesign existing ones, on some aspect of Canadian studies which they will subsequently offer as part of their regular teaching load. Any faculty member who holds the Ph.D. (or equivalent) from an accredited institution and is teaching courses related to Canada, or wishes to do so, is eligible. Candidates must have held a full-time teaching position at their university for at least two years. Written commitment from his or her university administration indicating that the applicant will offer this course at least three times within six years following completion of the period of study is required. A monthly stipend up to $1,500 for a period of no more than six months may be paid. The project must be undertaken between April 1, 1985 and February 1, 1986. The deadline for application is October 31, 1984. For details, contact Dr. Norman T. London, Faculty Enrichment Programme (see address below).

The Senior Fellowship in Canadian Studies is designed to facilitate the completion of a book or major monograph which would be of widespread interest to the Canadian studies community. Faculty members from four-year U.S. universities or colleges who wish to pursue a suitable project on a full-time basis may apply. A monthly stipend up to $3,000 for up to six months will be paid. There are no application forms. A letter of application outlining the proposed project and including a curriculum vitae, letters from three referees, and an estimated budget should be addressed to Dr. Norman T. London, Academic Relations Office, Public Affairs Division, Embassy of Canada, 1771 N Street, NW, Washington, D.C. 20036. (Telephone 202+785-1400, ext. 320) The same deadlines given above apply.

Development Exactions is the inaugural research project for the Homer Hoyt Center for Land Economics and Real Estate at Florida State University. In the 1945–77 period, Hoyt researched over 350 regional and local U.S. shopping centers. The collected analyses and research reports cover most aspects of land development. Microfiche or hard copies are available at cost of reproduction plus postage. Contact Dr. Edward A. Fernald, Homer Hoyt Center for Land Economics and Real Estate, 361 Bellamy Bldg, Florida State University, Tallahassee, FL 32306. (Telephone 904+644-2007)

Economists who are strongly oriented toward the humanities, who use humanistic methods in their research, and who will be participating in meetings held outside the United States, Mexico, and Canada that are concerned with the humanistic aspects of their discipline

are eligible to apply for small travel grants of the American Council of Learned Societies. Financial assistance is limited to air far between major commercial airports and will not exceed one-half of projected economy-class fare. Social scientists and legal scholars who specialize in the history or philosophy of their disciplines are eligible if the meeting they wish to attend is so oriented. Applicants must hold a Ph.D. degree or its equivalent, and must be citizens or permanent residents of the United States. To be eligible, proposed meetings must be broadly international in sponsorship or participation, or both. The deadlines for application to be received in the ACLS office are: meetings scheduled between July and October, March 1; for meetings scheduled between November and February, July 1; for meetings scheduled between March and June, November 1. Please request application forms by writing directly to the ACLS (Attention: Travel Grant Program), 800 Third Avenue, New York, NY 10022, setting forth the name, dates, place, and sponsorship of the meeting, as well as a brief statement describing the nature of your proposed role in the meeting.

Deaths

Ronald L. Teigen, professor of economics, University of Michigan, March 21, 1984.

Retirements

Martin Bronfenbrenner, William R. Kenan Jr. professor of economics, Duke University, May 1984.

H. Gregg Lewis, professor of economics, Duke University, May 1984.

Promotions

M. Akbar Akhtar: assistant director of research, Research and Statistics Function, Federal Reserve Bank of New York, January 1, 1984.

Bala Batavia: professor of economics, DePaul University.

M. William Dugger: professor of economics, DePaul University.

A. Steven Englander: chief, Business Conditions Division, Domestic Research Department, Federal Reserve Bank of New York, December 29, 1983.

Aaron S. Gurwitz: research officer and senior economist, Research and Statistics Function, Federal Reserve Bank of New York, January 1, 1984.

Helen Manning Hunter: professor of economics, Bryn Mawr College, September 1984.

Charles M. Lucas: vice president, Foreign Exchange Function, Federal Reserve Bank of New York, January 1, 1984.

Edward T. Merkel: associate professor of economics, Troy State University, fall 1984.

Carl J. Palash: research officer and senior economist, Research and Statistics Function, Federal Reserve Bank of New York, January 1, 1984.

Michael J. Piette: associate professor of economics, University of Hartford, March 1, 1984.

Beth Preiss: manager, Securities Research, Federal Home Loan Mortgage Corporation, January 1984.

Edward O. Price III: associate professor of economics, Oklahoma State University.

David L. Roberts: research officer and senior economist, Research and Statistics Function, Federal Bank of New York, January 1, 1984.

Andrew Silver: chief, Domestic Financial Markets Division, Financial Markets Department, Federal Reserve Bank of New York, December 29, 1983.

M. Raquibuz Zaman: professor of finance and applied economics, Ithaca College, February 24, 1984.

Administrative Appointments

John Laitner: associate chairman, department of economics, University of Michigan, July 1, 1984.

Roger Skurski: associate dean, College of Arts and Letters, University of Notre Dame, August 1983.

Ashton I. Veramallay: director, Center for Economic Education, Indiana University East, July 1, 1984.

Appointments

Soren Blomquist, University of Stockholm: adjunct associate professor, University of Michigan, July 1, 1984.

Andrew S. Carron, The Brookings Institution: vice president, Lehman Brothers Kuhn Loeb, Inc., March 20, 1984.

Suzanne Heller Clain: assistant professor of economics, Bryn Mawr College, September 1983.

A. W. Coats, University of Nottingham and National Center for the Humanities: research professor of economics, Duke University, January 1985.

Randall Crane, Massachusetts Institute of Technology: instructor of economics, Hollins College, September 1983.

Kevin Currier, State University of New York-Albany: assistant professor, Oklahoma State University, August 1984.

H. Randi DeWitty: economist, International Financial Markets Division, Financial Markets Department, Federal Reserve Bank of New York, January 11, 1984.

Timothy A. Deyak, Louisiana State University: associate professor of economics, University of South Florida, August 15, 1984.

Gertrud Fremling, University of Michigan: assistant professor of economics, University of Houston, September 1, 1984.

S. Malcolm Gillis, Harvard University: professor of public policy studies and economics, Duke University, September 1984.

Roger Gordon, Bell Laboratories: associate professor of economics, University of Michigan, January 1, 1984.

A. Ray Grimes, Oklahoma State University: director of research–regional economics, Rice Center, Houston, TX, March 1, 1984.

Yoo Soo Hong, Northwestern University: assistant professor, Oklahoma State University, August 1984.

Dalya Inhaber: economist, Monetary Analysis Division, Monetary Research Department, Federal Reserve Bank of New York, November 9, 1983.

Leslie A. Laufer, Yale University: visiting assistant professor of economics, Duke University, 1984–85.

Jeffrey Miron, Massachusetts Institute of Technology: assistant professor of economics, University of Michigan, September 1, 1984.

Robert D. Plotnick, Dartmouth College: associate professor of public affairs and social work, University of Washington, September 1984.

Partha Sen, London School of Economics: visiting assistant professor of economics, University of Michigan, September 1, 1984.

Ashton I. Veramallay: associate professor of economics, Indiana University East, July 1, 1984.

L. B. Wallerstein: instructor of economics, Montgomery College, January 1984.

Michelle White, New York University: professor of economics, University of Michigan, January 1, 1984.

John Yinger, Harvard University: visiting associate professor of economics, University of Michigan, September 1, 1984.

Leaves for Special Appointments

Felipe G. Morande, University of Houston: visiting professor, University of Santiago, Chile, fall and spring 1984–85.

Michael J. Piette, University of Hartford: senior research economist, Economic Research Services, Inc., Tallahassee, FL, July 1984–July 1985.

Frank A. Sloan: centennial professor of economics, Vanderbilt University, September 1984.

NOTE TO DEPARTMENTAL SECRETARIES AND EXECUTIVE OFFICERS

When sending information to the *Review* for inclusion in the Notes Section, please use the following style:

A. Please use the following categories:

1—Deaths
2—Retirements
3—Foreign Scholars (visiting the USA or Canada)
4—Promotions
5—Administrative Appointments
6—New Appointments
7—Leaves for Special Appointments (NOT Sabbaticals)
8—Resignations
9—Miscellaneous

B. Please give the name of the individual (SMITH, Jane W.), her present place of employment or enrollment: her new title (if any), new institution and the date at which the change will occur.

C. Type each item on a separate 3×5 card and please do not send public relations releases.

D. The closing dates for each issue are as follows: *March*, October 15; *June*, January 15; *September*, April 15; *December*, July 15.

All items and information should be sent to the Assistant Production Editor, *American Economic Review*, Room 8279, Bunche Hall, University of California, Los Angeles, CA 90024.

An Exceptional One-Term Economics Text . . .

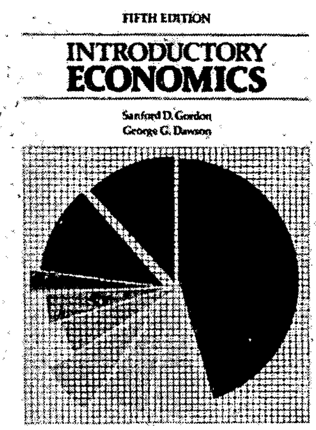
Available in hardback and paperback!

INTRODUCTORY ECONOMICS
Fifth Edition

Sanford D. Gordon, Russell Sage College, New York State Council on Economic Education

George G. Dawson, Empire State College, State University of New York

1984 Cloth 458 pages
1984 Paper 458 pages
Study Guide/Instructor's Guide



Concise and thorough, this highly successful introductory economics text uses an analytical/problem-solving approach to consider practical, real-world problems. Four review models — demand and supply, long-run equilibrium, pricing factors in production, and national income determination — expose students to micro-macro economic theory through practical applications to current problems and issues. Graphic and mathematical analyses presented in clear, simple steps are part of an effective pedagogical system that includes chapter objectives, summaries, highlights, key terms and names, review questions and problems, and a glossary of terms. Updated to reflect the economic changes of the 1980's, the Fifth Edition is ideal for short courses.

Highlights

- Recent developments in labor relations, including "givebacks," and growing interest in quality circles and other forms of Japanese workstyles.
- A careful examination of supply-side economics and a contrast with the demand-side approach of Keynes
- An updating of environmental problems and energy
- An explanation of the dramatic changes in our money and banking systems
- New information on our relations with the Common Market; the role of the dollar in the world economy; "Third World" difficulty in paying its external debt
- An appendix on economics in careers

From D. C. Heath!

*For details or sample copies, call us toll free: **800-225-1388.***
*In Massachusetts, call collect: **617-863-1948.***

A Raytheon Company

D. C. HEATH AND COMPANY

HEATH

College Division
125 Spring Street
Lexington, MA 02173

Fiscal and Monetary Policies and Problems in Developing Countries
ÉPRIME ESHAG

This book examines the use of fiscal and monetary policies to overcome three major obstacles faced by less-developed countries: inadequate investment, misallocation of investment resources, and internal and external imbalances (i.e., inflation and balance of payments deficits). Cloth $39.50 Paper $14.95

Applied General Equilibrium Analysis
HERBERT E. SCARF and JOHN B. SHOVEN, Editors

A collection of articles on applied general equilibrium analysis by some of the world's leading experts in this area of research, now involving the use of computers to simulate entire economies and the interrelationships among firms, households, and governments within them. $59.50

Theories of Inflation
HELMUT FRISCH

A survey of the new theories of inflation developed in the last two decades in response to the inflationary pressures experienced by the Western countries. Topics covered include the quantity theory, the inflation gap model, the foundations, stages, and modifications of the Phillips curve, monetarist inflation theory, and "stagflation". Cloth $39.50 Paper $12.50

Money: In Disequilibrium
DOUGLAS GALE

This wide-ranging study of the macroeconomic side of monetary theory deals with fundamental problems in existing theory and breaks new ground in the analysis of macroeconomic phenomena. The author takes a fresh look at such standard topics as wealth effects, money and growth, and the long-run effects of monetary and fiscal policy. Cloth $39.50 Paper $14.95

The Pocket Economist
RUPERT PENNANT-REA and WILLIAM EMMOTT

An alphabetically-guided tour of economics, economic institutions, and the world of high finance. ". . . a thoroughly practical and mercifully concise mix of encyclopedia and commentary, clearly and wittily written. Pennant-Rea is economics editor and Emmott a correspondent for *The Economist*, the best of news magazines. The quality shows."—*Newsday* $15.95

Cambridge University Press

Monetary Politics
The Federal Reserve and the Politics of Monetary Policy
JOHN T. WOOLLEY

This is the first book to describe and analyze the relationships between the Federal Reserve and the President, the Congress, bankers, and economists. John Woolley shows that, far from being politically independent, the Federal Reserve is sensitive to a wide range of political influences. $37.50

Now in paperback . . .

Growth, Profits, and Property
Essays in the Revival of Political Economy
EDWARD J. NELL, Editor

". . . excellent . . . the contributions are competent, stimulating, and relevant . . ."—*Journal of Economic Literature*
". . . it is a very welcome book, and its blend of constructive and destructive criticism of the neoclassical view of capitalism is generally quite effective . . ."—*The Economic Journal* Paper $16.95

Agriculture in China's Modern Economic Development
NICHOLAS R. LARDY

An exploration of the relationship between the Chinese peasantry, who are the fundamental base of support for the revolutionary Chinese Communist Party, and the state-led economic system established by the Party after 1949. The author seeks to explain how state policies have affected the growth of agricultural output, productivity, and the rise in peasant income and consumption, and asks how effectively agriculture has supplied food for urban consumption and export earnings and contributed to capital formation and the rise of modern industry. $37.50

Free Banking in Britain
Theory, Experience, and Debate, 1800–1845
LAWRENCE H. WHITE

The question of free banking as it confronted policy-makers and economic writers in Britain during the first half of the nineteenth century is explored in a study that encompasses monetary theory, economic history, and the history of economic doctrines. $29.95

Order from your bookstore or call our Customer Service department 1-800-431-1580 (outside New York State and Canada). MasterCard or Visa accepted.

Cambridge University Press • 32 East 57th Street • New York, N.Y. 10022

Environmental Policy under Reagan's Executive Order

The Role of Benefit-Cost Analysis

Edited by V. Kerry Smith

For the first time, a formal benefit-cost requirement plays an integral role in U.S. environmental policymaking. In this study of President Reagan's Executive Order No. 12291, some of the nation's leading experts on environmental policy appraise the effects of the order. By considering how the Environmental Protection Agency has responded to 12291, these essays identify the limitations of conventional practices of benefit-cost analysis.

Contributors are Richard N. L. Andrews, Ann Fisher, Arthur G. Fraas, W. Norton Grubb, Dale Whittington, Michael Humphries, A. Myrick Freeman III, Carlisle Ford Runge, Robert W. Crandall, and Paul R. Portney.

Approx. 300 pp., $25
Urban and Regional Policy and Development Studies

The University of North Carolina Press **PO Box 2288, Chapel Hill, NC 27514**

From the IMF. . .

New Occasional Papers

Three studies by the staff of the Fund's Research Department examine various aspects of the exchange rate system:

No. 28. *Exchange Rate Volatility and World Trade*

No. 29. *Issues in the Assessment of the Exchange Rates of Industrial Countries*

No. 30. *The Exchange Rate System: Lessons of the Past and Options for the Future*

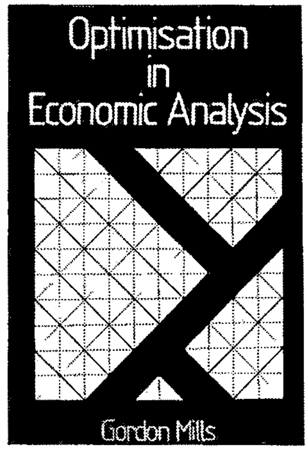
Price: $7.50 each ($5.00 each for university libraries, faculty, and students)

Advice on payment in currencies other than the U.S. dollar will be provided on request. For further information and to place orders, please write to:

Publications Unit, International Monetary Fund
Box E-164, Washington, D.C. 20431, U.S.A.

ALLEN & UNWIN

OPTIMISATION IN ECONOMIC ANALYSIS

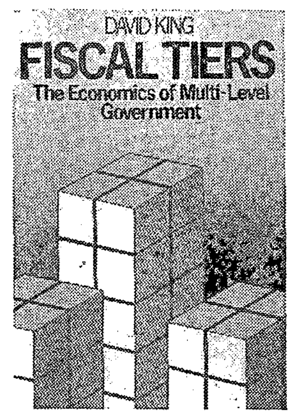


Gordon Mills

Here is a genuinely introductory textbook providing comprehensive and in-depth coverage of the use of mathematical optimization techniques in applied and policy contexts, as well as in economic theory. The book deals not only with the main techniques but also emphasizes model-building and discusses related areas such as optimization over time.

Paper $10.95 Cloth $27.50 208pp., illus. 1984

FISCAL TIERS

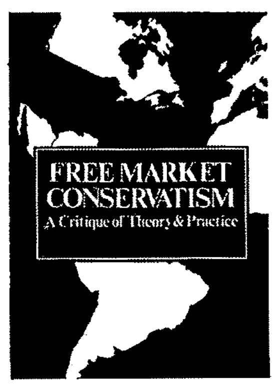


The Economics of Multi-Level Government
David King

This valuable new work brings together and develops the aspects of economics which relate to the design and operation of systems of non-central government. It examines the functions best served by these systems, optimal structures and sizes of their agencies, and the various means through which they are financed.

Paper $12.95 Cloth $29.50 320pp. 1984

FREE MARKET CONSERVATISM



A Critique of Theory and Practice
Edited by Edward J. Nell

This impressive collection of essays argues that conservative economics — be it supply-side, monetarism, or rational expectations — is inadequate in theory and often disastrous in practice. Reagan's U.S., Thatcher's U.K., and Pinochet's Chile are examined in turn, followed by an analysis of economic conservatism's ideology and social policy, and the reasons for its continuing appeal.

Paper $10.95 Cloth $27.50 256pp., illus. 1984

Fifty Cross Street
Winchester, MA 01890

For MC/VISA Orders Call Toll Free
1-800-547-8889

DEFY THE LAWS OF ECONOMETRICS.

Somewhere it is written that large-scale econometric applications require large computers. That the necessary speed and accuracy are available only on a mainframe.

It is also accepted that econometric software must be difficult to learn and complex to use. And that sizeable monthly time-sharing fees are a simple fact of life.

With all due respect to these principles of the past, we proudly present what's new: ESP™ The Econometric Software Package™ from Alpha Software® It's a full-scale, fast and highly accurate IBM® PC and XT version of the well-known ESP mainframe program.

And it's already been selected for use by Chase Econometrics, Prudential-Bache and New York Telephone, among others.

Why? Because nothing of the mainframe program is sacrificed. You get the analytical tools you need for sophisticated modeling and forecasting. Nearly all accepted statistical and regression techniques are at your command in one truly powerful program.

In fact, ESP from Alpha is actually enhanced with features like built-in graphics and advanced data management. Plus, ESP interacts like the best personal computer software—with a spontaneous, natural feel that's instinctive, and easy to grasp. It even has on-line help and tutorials.

ESP lets you download from any major data bank including Chase and DRI. Or move your own data from other PC software like 1-2-3® Multiplan® dBASE II® and our own Data Base Manager II™

But what's best is ESP gives you nearly instant results. Any analysis is performed before your eyes. Models, forecasts, scenarios—as many as you'd like, as often as you'd like, any time you'd like. At no incremental cost.

ESP™
The Econometric Software Package™

And you pay just $795 for ESP from Alpha.

For more information and a demonstration version of ESP, send $5.00 to Alpha Software Corporation, 30 B Street, Burlington, MA 01803. Or call us at 1-800-451-1018 (in Massachusetts call 1-800-462-2016).

ALPHA®
SOFTWARE CORPORATION

ESP and the Econometric Software Package are trademarks licensed to Alpha Software Corp. by MIKROS Corp. IBM is a registered trademark of International Business Machines Corp., 1-2-3 of Lotus, Multiplan of Microsoft, and dBASE II of Ashton-Tate. Data Base Manager II is a trademark of Alpha Software Corp. Alpha Software Corp. is a registered trademark. © 1984 Alpha Software Corp.

HYPERINFLATION OR DEPRESSION?
by Richard G. Zambell

Given the extreme duress presently surfacing in the U.S. and international financial systems, this timely book shows why traditional business cycle analysis is inadequate if one is trying to accurately forecast the economy.

It will serve as an excellent supplement in courses on business cycle forecasting, monetary theory, and macroeconomic theory. In the process of explaining the present weakness of business cycle analysis techniques, Mr. Zambell details the explicit interrelationships between the real and financial sectors of the economy. The text shows how monetary and macroeconomic theories are applied to the everyday world to impact economic growth and inflation.

A major point of the work is that each of the traditional economic theories (i.e. Keynesian, monetarist, supply-side, and rational expectations) has its own time and place during the much longer 40-60 long-wave cycle. Rather than one theory operating alone—at all times—it is more appropriate to synthesize these theories within the long-wave analysis if one is to correctly forecast and detail the economy's inner workings.

Mr. Zambell's previous work entitled *Linked Weekly, Monthly, Quarterly Econometric Models of the U.S. Economy*, provides an explicit mathematical foundation for the views expressed in *Hyperinflation or Depression?*

Richard G. Zambell, formerly Chief Economist at a major bank, is currently Director of Economic Research at Weiss Research, Inc.

CONTENTS

An Overview: The Causes and Consequences of the 1980s Depression
Money Growth Sets the Pace
The Role of the Stock Market
The 40-60 Year Long Wave
The 150-Year World Supremacy Cycle
The Outlook for the 1980s
The 21st Century

How the Key Decision-Makers Are Losing Control
The Federal Reserve
The Real Causes of Inflation and Recession
Summary: How the Economy Operates
The Fed's Catch-22 Dilemma

The Long Wave
The First Decade—Recovery
The Second Decade—Wealth Accumulation and War
The Fourth Decade—Transition and Stagnation
The Fifth Decade—Depression
Consequenses of Disbelieving the Long Wave
Why the Depression Can't Be Prevented
The Fallacy of the Keynesian Solution
The Fallacy of the Supply-Sider Solution
The Fallacy of the Monetarist Solution

The World Supremacy Cycle
The Regularity of Wars
The Regularity of Depressions
Depressions of the First Long Wave
Depressions of the Second Long Wave
Depressions of the Third Long Wave
What It Takes to Be the World's Supreme Power

The 1980s Depression-Four Scenarios
Four Scenarios
Most Likely Case—International Financial Collapse
The First Alternative—Double Digit Inflation
The Second Alternative—Roller Coaster Depression
The Third Alternative—Rampant Inflation
Why Hyperinflation is Unlikely in the 1980s
Forecast Summary

How to Escape the 1980s Financially Intact
Advice for Individuals
Advice for Businesses

Will the U.S. Lose World Supremacy in the 21st Century?
The Fate of Previous World Leaders
Who's Next in Line?
The Threat of Hyperinflation in the 21st Century
How to Retain World Leadership
Summary and Epilogue

Appendix

Notes

Bibliography

164 Pgs. $19.95

WR **Weiss Research, Inc.**
P.O. Box 2923
West Palm Beach, FL 33402

Enclosed is $ ________ ($19.95 per copy) for ________ copies of **Hyperinflation or Depression?**

Name ____________________

Address ____________________

City & State ____________________ Zip ________

All orders from outside North America must be paid in U.S. dollars by International Money Order only. Foreign orders—add $2 for each book via surface mail. Via air: all countries, add $10 for the first book and $7 for each additional. Canadian orders, add 25% if paying with Canadian check/currency.

MAIL THIS COUPON OR CALL (305) 684-8100

AEA sponsored Group Life Insurance for you and your family— at attractive rates!

The AEA Group Life Insurance Plan can help provide valuable supplementary protection—at attractive rates—for eligible members and their dependents.

Because AEA participates in a large Insurance Trust which includes other scientific and technical organizations, the low cost may be even further reduced by premium credits. In the past seven years, insured members received credits on their April 1 semiannual payment notices averaging over 44% of their annual premium contributions. (These credits are based on the amount paid during the previous policy year ending September 30.) Of course future premium credits, and their amounts, cannot be promised or guaranteed.

Now may be a good time for you to re-evaluate your present coverage and look into AEA Life Insurance. Just fill out and return the coupon for more details at no obligation.

Administrator, AEA Group Insurance Program F-6
1255 23rd Street, N.W.
Washington, D.C. 20037

Please send me more information about the AEA Life Insurance Plan.

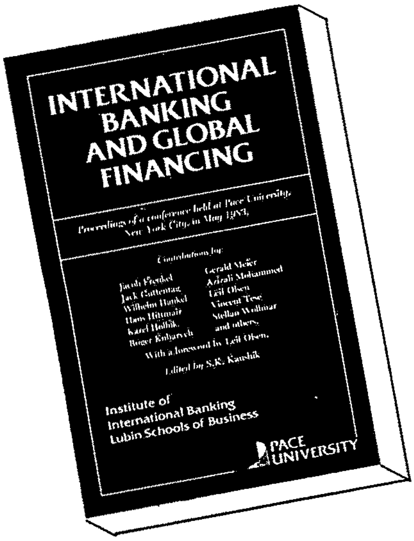
Name ______________________ Age ________

Address ______________________

City ______________ State ________ Zip ________

Or—call today Toll-Free 800-424-9883
(Washington, DC area, call 296-8030)

Please mention THE AMERICAN ECONOMIC REVIEW *When Writing to Advertisers*



JUST PUBLISHED

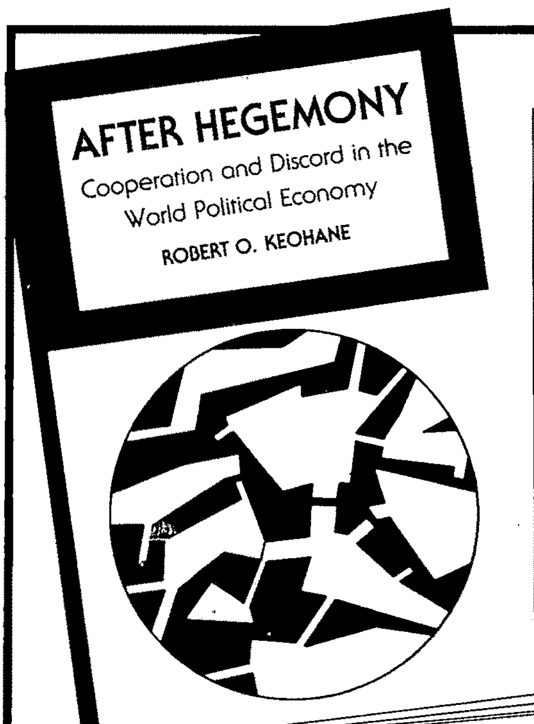
CONTRIBUTORS:

Jacob A. Frenkel: "International Liquidity and Monetary Control". **Jack Guttentag & Richard Herring:** "What Happens When Countries Cannot Repay Their Bank Loans? The Renegotiation Process". **Gerald M. Meier:** "Analyzing the Risk of Lending to Developing Countries". **Leif H. Olsen:** "The Bretton Woods Nostalgia". **Wilhelm Hankel:** "The Global Financial Crisis — A Case Study in Modern Alchemy — Facts, Legends and Realities". **Karel Holbik:** "International Interdependence, Financial Intermediation and Multinational Banking". **And others.**

Edited by S.K. Kaushik

INTERNATIONAL BANKING AND GLOBAL FINANCING

257 pages, paperback. $16. postpaid. To order, send check or money order, payable to **Pace University,** to: Pace University, Attn: Comptroller, Pace Plaza, New York, NY 10038.



"A brilliant piece of scholarship—erudite and refined, as well as clearly organized and written. Keohane advances our understanding both of the factors that give rise to the creation and adaptation of international regimes and of the various roles that regimes play. His book will be a standard in the field for years to come."
—*Benjamin Cohen, Tufts University*

P $8.95. C $30.00

Available from

Princeton University Press 41 William Street, Princeton, NJ 08540

New from MIT

Multinational Excursions

Charles P. Kindleberger

"This new collection of Kindlebergeriana rescues lucid pieces from obscure outlets, and makes available to all lectures only a lucky few have heard. Aficionados of multinationals, scholars, teachers, and other Kindleberger groupies will delight in this fine volume."
–Carlos F. Diaz-Alejandro, Columbia University

These papers, lectures, articles, and reviews, prepared over the past decade focus on the role of multinational corporations in the international economy, their relationships with home and host countries, the determinants of their size, the impetus to their investment behavior, their history, the literature about them, and their regulation.
$22.50

The Future of the Automobile

The Report of MIT's International Automobile Program
by Alan Altshuler, Martin Anderson, Daniel Jones, Daniel Roos, and James Womack

A new shape for the world automobile industry emerges from this seven nation, four-year study, the most comprehensive assessment ever conducted of the world's largest industry.
$16.95

Beyond Mechanization

Work and Technology in a Post-Industrial Age
Larry Hirschhorn

"*Beyond Mechanization* is a major work with enormous implications for current debates about industrial and economic policy. . . . It persuasively describes the emergent postindustrial workplace through a careful examination of some of the most advanced experiments in work redesign. In a relatively small volume–tightly argued, crammed with data, and highly readable–Hirschhorn has produced a classic work that forces us to see the world differently."– Fred Block, Associate Professor of Sociology, University of Pennsylvania
$17.50

Deregulation and the New Airline Entrepreneurs

John R. Meyer and Clinton V. Oster, Jr.

Deregulation in the airline industry has spurred a major growth in entrepreneurs competing with established carriers. This book takes a close look at commuter airlines using small propeller-driven aircraft and jet-equipped carriers like the now familiar New York Air and People's Express, focusing in particular on how these two relatively new segments of the industry will eventually be structured, and their impact on the rest of the industry.
Regulation of Economic Activities Series, edited by Richard Schmalensee.
$22.50

Write for our current economics catalog

28 Carleton Street Cambridge, MA 02142 **THE MIT PRESS**

Recent Best Sellers

Economics for Policymaking

Selected Essays of Arthur M. Okun

edited by Joseph A. Pechman

"Arthur Okun was the wisest and most creative economic policy advisor of our times. His posthumous book is required and enjoyable reading for non-economists as well as professionals."
–Paul Samuelson
$25.00

Money and Inflation

Frank Hahn

This brilliant critique demonstrates in unequivocal terms that monetarism offers an implausible solution to the most pervasive economic problems.
$13.75

Inflation, Debt, and Indexation

edited by Rudiger Dornbusch and Mario Henrique Simonsen

Indexation has been a much proposed and debated tool of economic policy, and the interaction of inflation, private contracts and pricing, and public sector financing constitutes an important and lively issue in monetary economics. These essays cover frontier topics in modern macroeconomics and macropolicy, ranging from the highly theoretical to practical experiences in Brazil and in Europe.
$25.00

Folded, Spindled, and Mutilated

Economic Analysis and *U.S.* v. *IBM*

Franklin M. Fisher, John J. McGowan and Joen E. Greenwood

"This is an invaluable book, laying out in intelligible fashion the economics of recent cases under Section II of the Sherman Act. No student of industrial organization or antitrust policy can afford to ignore this volume, which clearly lays out the pertinent analysis and shows, using the IBM case as its illustration, the perils of ignoring its logic."–William Baumol, Princeton University
$27.50

Write for our current economics catalog

28 Carleton Street Cambridge, MA 02142 **THE MIT PRESS**

TEACHING MARXISM

Works by Bertell Ollman, Prof. of Politics, NYU

1. **Left Academy: Marxist Scholarship on American Campuses,** vol. II, co-ed with Edward Vernoff (Praeger, 1984). Disciplines covered: Literature, Education, Law, Geography, Classics, Art History, and Biology.
2. **Left Academy: Marxist Scholarship on American Campuses,** vol. I, co-ed with Edward Vernoff (McGraw Hill, 1982). Disciplines covered: Political Science, Sociology, Economics, Philosophy, History, Psychology, and Anthropology.
3. **Alienation: Marx's Conception of Man in Capitalist Society** (Cambridge U.P., 1971 and 1976, 7th printing). Adopted in classes at over 300 universities. "A remarkable book . . . Ollman has discovered a kind of Ariadne's thread through the maze of Marxian linguistic problems . . . brilliant and illuminating". New York Review of Books.
4. **Social and Sexual Revolution: Essays on Marx and Reich** (South End Press, 1979). Emphasis is on class consciousness, dialectical method, and Marx's vision of communism.
5. **Studies in Socialist Pedagogy,** co-ed with Ted Norton (Monthly Review Press, 1978). How-to-do-it book for (and by) socialist teachers.
6. **"What is Marxism? A Bird's-Eye View",** (Red Hot Publications, 1982). A brief (10pp.), clear, humorous overview that touches all the main bases. For introductory classes or study groups.
7. **Class Struggle** board game, (Avalon Hill Game Co., 4517 Harford Rd., Baltimore, Md., 1983). New book case edition of first Marxist board game. Has been used in over 100 college and high school class rooms to simulate the real thing. "More fun than Das Kapital". Paul Sweezy and Harry Magdoff.
8. **Class Struggle** computer game (for Apple II computer only at this time), (Avalon Hill Game Co., 1984). First Marxist computer game. Now you can also assign it for home playing/reading.
9. **Class Struggle Is the Name of the Game: True Confessions of a Marxist Businessman** (Wm. Morrow Pub., 1983). Like it says. Also contains an account of the author's academic freedom controversy with the Univ. of Maryland. "In these pages, Karl and Groucho Marx finally meet and share their respective talents for critical economic theory and madcap humor. A must text for micro-economic courses". James O'Connor.

(For bulk orders and trade discounts, contact publisher directly;
for individual orders and bulk orders of pamphlet, see below.)

		No.			No.
1. **Left Academy,** II (pb)	– **$10.95**	______	6. **"What is Marxism?"** (Min. Order of 10 @ 50¢ ea.)	– **$ 5.00**	______
2. **Left Academy,** I (pb)	– **$ 8.95**	______			
3. **Alienation** (pb)	– **$11.95**	______	7. **C.S.** board game	– **$17.60**	______
4. **Soc. and Sex. Rev.** (pb)	– **$ 5.50**	______	8. **C.S.** comput. game	– **$33.00**	______
5. **Soc. Pedagogy** (pb)	– **$ 6.50**	______	9. **True Confessions** (hb)	– **$12.95**	______

Please add $1.50 for first work and $.50 for each subsequent work (bulk orders of the pamphlet count as a single work) to cover handling and mailing costs. Add state and local tax if applicable. Make out check or money order to **Class Struggle, Inc.,** and send to **Class Struggle Inc., 4 Washington Square Village (9A), New York, NY 10012.**

Chicago

Money, Capital, and Fluctuations

Early Essays

F. A. Hayek

Edited by **Roy McCloughry**

Friedrich Hayek, winner of the 1974 Nobel Prize in Economics and author of *The Road to Serfdom,* is one of the best-known economists of our day. This volume presents the first English translations of some of the most important articles from the period "when," writes Hayek, "I was slowly developing my own distinct interests in economic theory but had not yet clearly perceived which of them were to become my chief concerns in the field."

Cloth $20.00   200 pages

The Economist as Preacher

And Other Essays

George J. Stigler

In this collection of essays, Nobel Prize winner George J. Stigler — renowned for his urbane good humor and mischievous wit — looks to the history of economics to examine the role of economists as advisors and the belief that economists have been influential in the devising and execution of public policy. "[*The Economist as Preacher*] brings together the intellectual fruits of a distinguished career, and forms a surprisingly unified whole." — Robert Dorfman, *Journal of Economic Literature*

Paper $8.95   268 pages

Great Economists

1985 Engagement Calendar

Compiled by **George J. Stigler** and **Claire Friedland**

This unique engagement calendar presents a selection of fifty-two economists, recent and ancient, from towering geniuses to the merely important. All are represented by portraits, brief extracts from their writings, and, in many cases, comments about them — witty, malicious, or simply true — by other economists.

Spiral bound $9.95
128 pages

The University of Chicago Press

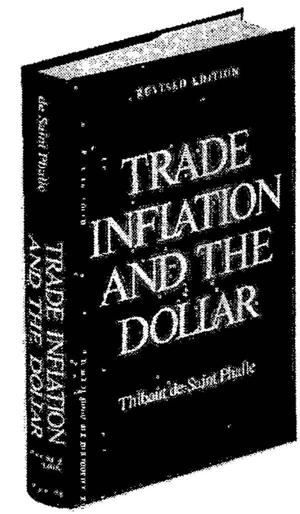
5801 South Ellis Avenue   Chicago, IL 60637

Important Trends in Economic Theory and Practice...

Completely Updated Facts and Figures. . .

TRADE, INFLATION AND THE DOLLAR

Revised Edition
Thibaut de Saint Phalle



Exactly how has the global economy changed over the past 30 years and what does this mean to government officials, business executives, labor leaders and entrepreneurs? Anyone seeking basic knowledge of international economics or global business strategy will find the answers here. Among the topics discussed are the monetary system and the effect of the strong dollar on the decline in U.S. inflation; the inflows and outflows of currency which have balanced the increased trade deficits of the United States; and the role of energy imports to the trade balance.

464 pp.	July 1984	$38.95	ISBN 0-03-070773-0
Paper edition		$14.95	ISBN 0-03-069552-X

FORECASTING WITH COMPUTER MODELS

Econometric, Population and Energy Forecasting

Rodney D. Green

A methodical training guide for students and professionals in the use of computer models for economic and demographic policy analysis. Takes you step by step through various actual models now being used by government and private researchers.

ca. 288 pp.	October 1984	ca. $29.95	ISBN 0-03-063788-0

THE ECONOMICS OF THE TOBACCO INDUSTRY

Paul R. Johnson

172 pp.	April 1984	$27.95	ISBN 0-03-062562-9

AMERICA'S ELECTRIC UTILITIES

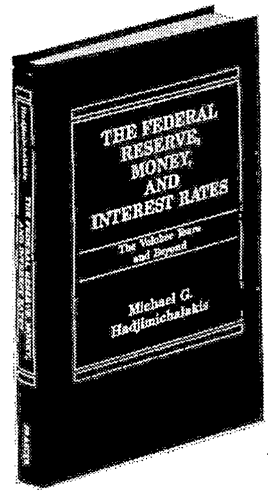
Under Siege and In Transition

Scott Fenn

176 pp.	May 1984	$23.95	ISBN 0-03-070301-8

PRAEGER PUBLISHERS 521 Fifth Avenue New York, N.Y. 10175

THE FEDERAL RESERVE, MONEY, AND INTEREST RATES



The Volcker Years and Beyond

Michael G. Hadjimichalakis

Analyzes the contemporary operations and goals of the Federal Reserve and examines analytically and factually the record of the Volcker years. Explains the errors that led to the 1981–82 recession of the emergence of the new financial environment of (financial) deregu·lation.

288 pp. June 1984 $29.95 ISBN 0-03-063861-5

NEW DIRECTIONS IN ECONOMIC POLICY

An Agenda for the 1980s

Everett M. Ehrlich and **Raymond C. Scheppach**

Here is a lively new book on domestic economic policy which defines and analyzes the major economic issues facing America during the decade of the '80s.

224 pp. June 1984 $27.95 ISBN 0-03-063786-4

U.S. WOOD-BASED INDUSTRY

Industrial Organization and Performance

Paul V. Ellefson and **Robert N. Stone**

ca. 320 pp. September 1984 ca. $25.95 ISBN 0-03-063698-1

GREECE AND YUGOSLAVIA

An Economic Comparison

Nicholas V. Gianaris

272 pp. June 1984 $34.95 ISBN 0-03-071466-4

INDUSTRIAL RELATIONS IN LATIN AMERICA

edited by **Efren Cordova**

288 pp. March 1984 $28.95 ISBN 0-03-070284-4

Available through your local bookseller, or order directly from:

PRAEGER PUBLISHERS **521 Fifth Avenue New York, N.Y. 10175**

The Text for Tomorrow's Decision-Makers...

MANAGERIAL ECONOMICS

BY STEVEN T. CALL
Metropolitan State College &
University of Colorado, Denver
& WILLIAM L. HOLAHAN
University of Wisconsin
at Milwaukee

Here is the text that gives students the conceptual framework to understand the decision-making process. How? Through models that explain actual economic behavior in the marketplace. Through a verbal/graphical approach that encourages conceptually sophisticated study (without losing students in peripheral math concerns). Through current applications at the proper level that link managerial economics with the real world and reinforce understanding of theory.

Watch for Call & Holahan's
PRINCIPLES OF ECONOMICS, coming in Fall 1985!

NEW FROM WADSWORTH...

INTERNATIONAL ECONOMICS, Second Edition
By Robert J. Carbaugh, University of Wisconsin/ Eau Claire

COMPARATIVE ECONOMIC SYSTEMS, Second Edition
By John E. Elliott, University of Southern California

THE ECONOMICS OF DEVELOPING COUNTRIES
By E. Wayne Nafziger, Kansas State University

ECONOMICS: PRINCIPLES IN ACTION, Fourth Edition
By Philip C. Starr, Chaffey College

BASIC STATISTICS IN BUSINESS AND ECONOMICS, Fourth Edition
By George W. Summers, Arizona State University; William S. Peters, University of New Mexico; and Charles P. Armstrong, University of Rhode Island

For additional information on these texts, write Stephanie Surfus, Economics Editor, Box AER-9-84, at the address below.

WADSWORTH PUBLISHING COMPANY
Ten Davis Drive • Belmont, California 94002

NORTH-HOLLAND ANNOUNCES

Introduction to the Use of General Equilibrium Analysis

by RICHARD R. CORNWALL, *Middlebury College, Middlebury, Vermont, U.S.A.*

ADVANCED TEXTBOOKS IN ECONOMICS, Volume 20

1984 xvi + 774 pages
US $72.50 (in USA & Canada)
Dfl. 200.00 (Rest of World)
ISBN 0-444-86443-1

In this book general equilibrium theory is presented as an approach which is useful for policy analysis. The foundations for constructing and using general equilibrium simulation models include fixed-point algorithms, relations between competitive market solutions and voluntary bargaining solutions (the core), description of efficient and fair solutions including possible externalities and the use of notions of consumer surplus and of Debreu's coefficient of resource utilization to evaluate how well an economy performs.
Students are offered a careful use of axiomatic method as well as a systematic survey of the mathematics needed for the advanced parts of this text.

Robustness of Bayesian Analyses

edited by JOSEPH B. KADANE

STUDIES IN BAYESIAN ECONOMETRICS, Volume 4

1984 xii + 310 pages
US $57.75 (in USA & Canada)
Dfl. 150.00 (Rest of World)
ISBN 0-444-86209-9

Robustness is a fundamental issue for all statistical analyses; in fact it might be argued that robustness is the subject of statistics. In Bayesian statistics, the prior distribution can be seen as weighting the possible values of the parameter by their probability. The studies reported in this volume concern the sensitivity of Bayesian analyses to the inputs: the prior, the likelihood, and, where applicable, the utility function. There are many approaches to bayesian robustness, some of which are represented, and many of which are mentioned in this volume.

Trade, Income Levels, and Dependence

by MICHAEL MICHAELY, *The Hebrew University of Jerusalem*

STUDIES IN INTERNATIONAL ECONOMICS, Volume 8

1984 xvi + 188 pages
US $38.50 (in USA & Canada)
Dfl. 100.00 (Rest of World)
ISBN 0-444-86771-6

This is a study of attributes of size and structure of foreign trade and their implications for several major issues of concern for national and international economic policies. The study is divided into two parts: 'Trade Dependence: Concepts and Measurements', and 'Income Levels and Trade Patterns'.
Each of the two parts of the study is constructed as an integrated and systematic discussion, within which the analyses of issues follow on from each other and depend on each other.

Price Expectations in Rising Inflation

by IGNAZIO VISCO, *Banco d'Italia*

CONTRIBUTIONS TO ECONOMIC ANALYSIS, Volume 152

1984 xii + 272 pages
US $46.25 (in USA & Canada)
Dfl. 120.00 (Rest of World)

It is claimed in this book that expectations should not necessarily be treated as unobservable variables and that there is much to be learned from survey data. A unique data set is examined, the output of surveys conducted twice a year since 1952 among informed Italian businessmen and economic experts. The predictive accuracy, rationality and determinants of inflation expectations are investigated, following an extensive analysis of measurement issues.

ELSEVIER SCIENCE PUBLISHERS
P.O. Box 211, 1000 AE Amsterdam, The Netherlands

For customers in the U.S.A. and Canada:
ELSEVIER SCIENCE PUBLISHING CO., INC.
P.O. Box 1663, Grand Central Station, New York, N.Y. 10163

NEW BOOKS · NEW BOOKS · NEW BOOKS · NEW

0886NH/BKS/ECON

Books that matter are Basic

The Global Debt Crisis

America's Growing Involvement

John H. Makin

The $700 billion global debt crisis now facing us is unique, it is serious, and it will not go away. This is the sobering conclusion of this important book, the first definitive account of how the current crisis started, and who is responsible. The author, a distinguished economist who has had the opportunity to observe the monetary crisis both from the outside and from within the International Monetary Fund, points out that more defaults are likely: much of the interest on the outstanding loans *cannot* be paid and the political climate is such that much of the principal *will not* be paid. How did we get into this mess—and how can we extricate ourselves? Writing with rare wit and clarity, Professor Makin explores the background of this growing crisis and explains its implications for future government policy.

October $18.95

The Second Industrial Divide

Michael J. Piore and Charles F. Sabel

Modern industrial society is inconceivable without the assembly line, the giant corporation, and mass production. Or is it? Two McArthur Prize Fellows present a brilliant argument that we abandon our attachment to standardized mass production and seek instead prosperity through a system of flexible specialization. The authors point to the emergence of such new systems of industrial production that make use of the latest advances in microelectronics and computers as well as of the more conventional technologies of the textile, construction, and other industries. Whether we are about to witness the birth of a new industrial order is by no means decided, but Piore and Sabel point out that we have a choice; a more efficient, flexible, and humane alternative exists. October $21.95

Of Politics and Economic Reality

The Art of Winning Elections with Sound Economic Policies

Amar Bhide

In this broad-ranging view of recent economic policy mistakes in the non-communist world, Amar Bhide concludes that the bad economic policies of our day have almost always resulted from poor political judgements. He shows how time and again politicians have implemented bad policies they thought were good policies—only to be promptly voted out of office. Bhide examines the *political* fallacies of Reaganism and Thatcherism, of supply siders, and of proponents of a vigorous industrial policy, and concludes that if only politicians abandoned such panaceas and concentrated instead on attracting votes in the next election we would be seeing more growth, lower unemployment, and stable prices. $15.95

Industrial Renaissance

Producing a Competitive Future for America

William J. Abernathy, Kim B. Clark, and Alan M. Kantrow

Three leading analysts affiliated with the Harvard Business School argue that the current crisis in American industry results less from external factors, like energy costs, than from the failure of American managers to adjust to the realities of today's global markets. In a book that has already stirred controversy, they explain how the very nature of competition has changed, what demands these changes place on management, and what promises they hold out for the future. "One of the best arguments yet that the blame for industry's decline in this country sits squarely on the lap of management."—*Wall Street Journal* Cloth $19.00 Paper $ 8.95

Basic Books, Inc.

10 East 53rd Street New York, N.Y. 10022

New Economic Research

Industrial Policies: International Restructuring and Transnationals
Jack N. Behrman, The University of North Carolina
In these brilliant essays, Behrman calls for global reordering of the economy in order to establish a fair balance of industrial benefits. He recommends intensifying government-business dialogue and cooperation and restructuring international industrial patterns in ways that could enhance industrial efficiency.
272 pages
ISBN 0-669-08275-9 $23.50

Oil Price Effects and OPEC's Pricing Policy
An Optimal Control Approach
Jaime R. Marquez, Board of Governors of the Federal Reserve System
An experienced economist produces an innovative, twofold analysis of issues central to the changing world economy—the determinants of oil price changes and their effects. He creates a world model of developed, OPEC, and non-OPEC developing economies to study these issues. Among his conclusions is that increased oil prices need not reduce real income in industrialized countries. *The Wharton Econometric Studies Series.*
August
ISBN 0-669-08200-7 ca. $34.00

Industrial Policies for Growth and Competitiveness
An Economic Perspective
edited by F. Gerard Adams and Lawrence R. Klein, University of Pennsylvania
A high-level examination of industrial policies at home and abroad that provides keen insight into how to develop remedial policies to stimulate U.S. industry. Representatives from international organizations, academic institutions, national governments, commodity organizations, and commodities businesses address the definitions, mechanisms, justifications, and criteria for industrial policies in developed and developing countries. *The Wharton Econometric Studies Series.*
448 pages
ISBN 0-669-05412-7 $34.00

Vanishing Farmland: *A Legal Solution for the States*

Sarah E. Redfield, Franklin Pierce Law Center
Redfield's model for state legislation to preserve agricultural land combines regulation with governmental benefits to farmers, all within the limits of constitutionality. She thoroughly assesses other past, current, and proposed legal techniques.
August
ISBN 0-669-08233-3 ca. $29.00

Lexington Books
125 Spring Street
Lexington, MA 02173
(617) 862-6650 (212) 924-6460
Call our toll-free numbers
(800) 428-8071 or (800) 33-HEATH

from Lexington Books

Oxford

Leading Issues in Economic Development
Fourth Edition

GERALD MEIER, *Stanford University.* Completely revised and updated, this new edition reflects the current development mood of stock-taking and reassessment. Bringing order to the diffuse subject of economic development without sacrificing a variety of viewpoints, the author makes readily accessible a meaningful range of theoretical, applied, and policy materials in emphasizing the strategic policy issues which accelerate the development of poor nations.
1984 768 pp.; 45 illus. paper $19.95

Taxation for Development
Principles and Applications

STEPHEN R. LEWIS, JR., *Williams College.* This introduction to the analysis of tax policy in developing countries focuses on the open economy and the real costs of subsidizing certain industries through tax policies. Sections on public finance, income taxation, wealth and property taxes, and indirect taxes examine the different kinds of taxes used in developing countries, emphasizing these with problem-solving cases.
1984 256 pp.; 39 illus. cloth $29.95 paper $12.95

Keynes's Economics and the Theory of Value and Distribution

Edited by JOHN EATWELL, *Cambridge University,* and MURRAY MILGATE, *University of Sydney.* This book exposes the controversy surrounding Keynes's theories and suggests an entirely new approach to the relationship between the theory of value and distribution and the analysis of output. The editors bring together articles from the *Cambridge Journal of Economics* and new works written specifically for the volume in examining the history, criticism, and analysis of Keynes's theories.
1983 304 pp. paper $9.95

Financing State And Local Government in the 1980s

ROY BAHL, *Syracuse University.* "An illuminating and comprehensive book... it contributes both to our stock of theory and to our views about the economic reality of states and localities. For the reader interested in a thorough instruction in both positive and normative issues concerning the economics of state and local governments, this book is recommended."—Gerasimos T. Soldatos, *Worcester State College*
1984 272 pp. cloth $24.95 paper $11.95

Microeconomics
Problems and Solutions

DAVID M. WINCH, *McMaster University.* This collection of problems for students in intermediate microeconomics provides a wide range of applications to the real world. Intended as a supplement to a standard text, it includes exercises and problems, hints at solutions, and worked-out solutions—helping the student to understand how microeconomic theory works in practice.
1984 225 pp. paper $9.95

Emerging from Poverty
The Economics that Really Matters

GERALD M. MEIER. *Advance comment:* "This is the story of both what happened in the developing countries, and how scholars have analyzed these events and policies. It conveys the excitement, as well as the disappointments of a great effort. The tension between the economist's role as trustee for the poor and as guardian of rationality, between the heart and the head, underlies this admirable story of the effort of raising the fate of 2 billion people from poverty."—Paul Streeten, *Director, World Development Institute, Boston*
1984 256 pp. $29.95

Pioneers in Development

Edited by GERALD M. MEIER and the late DUDLEY SEERS. The pioneers in development economics—Lord Bauer, Colin Clark, Albert O. Hirschman, Sir Arthur Lewis, Gunnar Myrdal, Raul Prebisch, Paul N. Rosenstein-Rodan, W.W. Rostow, H.W. Singer, and Jan Tinbergen—offer a fascinating retrospective view of their work, providing rare autobiographical detail and insight into why they said what they did, and what they think of the current state of development thought and policy.
1984 384 pp. $29.95

Money in Europe
Problems of Coordination and Integration

Edited by ROBERT TRIFFIN, *University of Louvain-la-Neuve,* and RAINER S. MASERA, *Banca d'Italia.* This record of the world monetary system over the last few decades highlights the major economic and political shortcomings that are largely responsible for world "stagflation" and recurrent exchange crises, and recommends a highly decentralized world monetary system that would fully exploit opportunities for mutually advantageous policies and negotiable commitments. *(Centre for European Policy Studies)*
1984 208 pp. $34.95

Industrial Policy Issues in Europe

Edited by ALEXIS JACQUEMIN, *Université Catholique de Louvain.* In this volume, articles by some of the leading experts on European industrial issues provide a critical and independent analysis of the main factual, methodological, and theoretical aspects of industrial policy.
(Centre for European Policy Studies)
1984 208 pp. $29.95

Europe's Stagflation
Causes and Cures

Edited by MICHAEL EMERSON, *Commission of the European Communities, Brussels.* Focusing on such topics as the interaction between external oil and monetary shocks with real wage rigidities and domestic financial policies, the role of the massive growth of the public sector in slowing Europe's economic potential, and the scope for new policy prescriptions, this study sheds new light on the fundamental nature of western Europe's "stagflation" and contributes to an understanding of the most serious issues currently faced by European economies.
(Centre for European Policy Studies)
1984 208 pp. $34.95

Prices are subject to change.

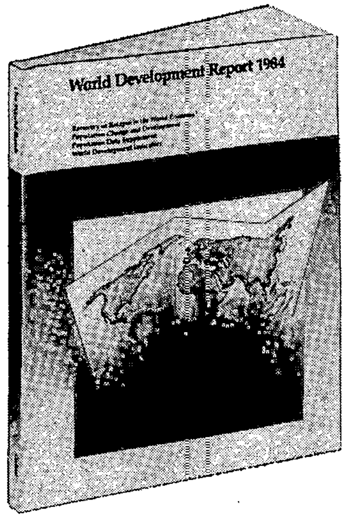
Oxford University Press
200 Madison Avenue • New York, New York 10016

Oxford

Now Available!

World Development Report 1984

THE WORLD BANK



The *World Development Report* is recognized worldwide as essential reading for economists, students, planners and policy makers, bankers and businessmen, journalists, and development professionals of all kinds. This year's edition of the *World Development Report* examines current signs of international economic recovery following years of recession. Opening with a review of trends in the international economy and their implications for the developing countries, the *Report* goes on to focus on population changes and their links with development. This year, the *Report's* regularly featured "World Development Indicators" section—more than 75 pages of the most up-to-date and comprehensive statistical information on social and economic development—is complemented by vital demographic information in an all-new "Population Data Supplement."

Up-to-date, authoritative information, valuable insights, and an abundance of supporting data make the *World Development Report* "the most useful single source on current global development trends." (John Adams, *University of Maryland*)

Contents

Recovery or Relapse in the World Economy?
Recession in Retrospect. Prospects for Sustained Growth

Population Change and Development
Demographic Change and Public Policy. The Consequences of Rapid Population Growth. Slowing Population Growth. Family Planning as a Service. The Policy Agenda. Ten Years of Experience

Population Data Supplement

Bibliographical Note

World Development Indicators

1984 286 pp. tables, charts, graphs, maps cloth $20.00 paper $8.00

Prices and publication dates are subject to change.

Oxford University Press
200 Madison Avenue • New York, NY 10016

Important new titles in economics

INTERNATIONAL COMPARISONS OF PRODUCTIVITY AND CAUSES OF THE SLOWDOWN

John W. Kendrick, Editor
An American Enterprise Institute book.
Informative essays that include a wealth of statistics on productivity trends and cycles, and offer the most comprehensive treatment yet of international productivity comparisons.
1984—448 pages—$39.95 cloth
ISBN 0-88410-971-2—LC 84-3107

NEOCLASSICAL POLITICAL ECONOMY
The Analysis of Rent-Seeking and DUP Activities

David C. Colander, Editor
Major economists present research that formally integrates political issues into economic theory. Some of the topics covered include: endogenous tariff theory, rent-seeking behavior, and monopolization. Diverse viewpoints from neoclassical institutionalist, public choice, new economic history, and property right theorists are all represented in this unique volume.
October 1984—ca 304 pages—$32.00 (tent.) cloth
ISBN 0-88410-999-2—LC 84-11124

MONEY IN CRISIS
The Federal Reserve, the Economy, and Monetary Reform

Barry N. Siegel, Editor
A Pacific Institute book.
A much-needed proposal for monetary reform that discusses Hayek's theory of economic fluctuations, capital formation in the United States, the Great Depression, Federal Reserve policy since 1945, the gold standard, free banking as an alternative monetary system, and other topics of importance to economists.
September 1984—ca 392 pages—$35.00 (tent.) cloth
ISBN 0-88410-962-3—LC 84-9230

THE LEVEL AND COMPOSITION OF HOUSEHOLD SAVING

Patric H. Hendershott, Editor
An American Council for Life Insurance book.
Analyzes the presumed decline in personal savings rates and the change in composition of these savings. The book covers government, business, and household savings, consumer durables, pension assets, human capital, and microeconomic studies of household saving behavior.
January 1985—ca 280 pages—$32.00 (tent.) cloth
ISBN 0-88410-987-9

FLOATING EXCHANGE RATES AND THE STATE OF WORLD TRADE AND PAYMENTS

David Bigman and **Teizo Taya**, Editors
An in-depth look by major economists at the turbulent international financial markets over the past decade. Explores in detail the implications for world trade, the effects on the stability of national economies, and the policy options of individual countries.
1984—336 pages—$15.00 paperback
ISBN 0-88410-998-4—LC 84-2905

EDUCATION AND ECONOMIC PRODUCTIVITY

Edwin Dean, Editor
What is education's contribution to the U.S. economy? This volume assesses economic growth and the equality of educational opportunity; existing ways of measuring the impact of education on productivity; education, organization design, and productivity; education, R & D and productivity rowth; and private and social returns on investment in education.
1984—240 pages—$29.95 cloth
ISBN 0-88410-943-7—LC 84-3046

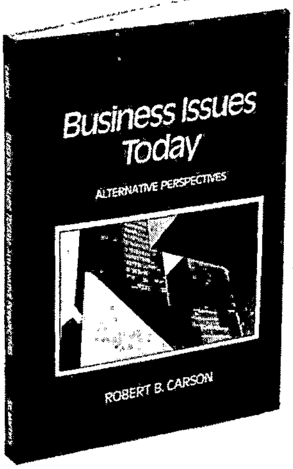
THE GLOBAL REPERCUSSIONS OF U.S. MONETARY AND FISCAL POLICY

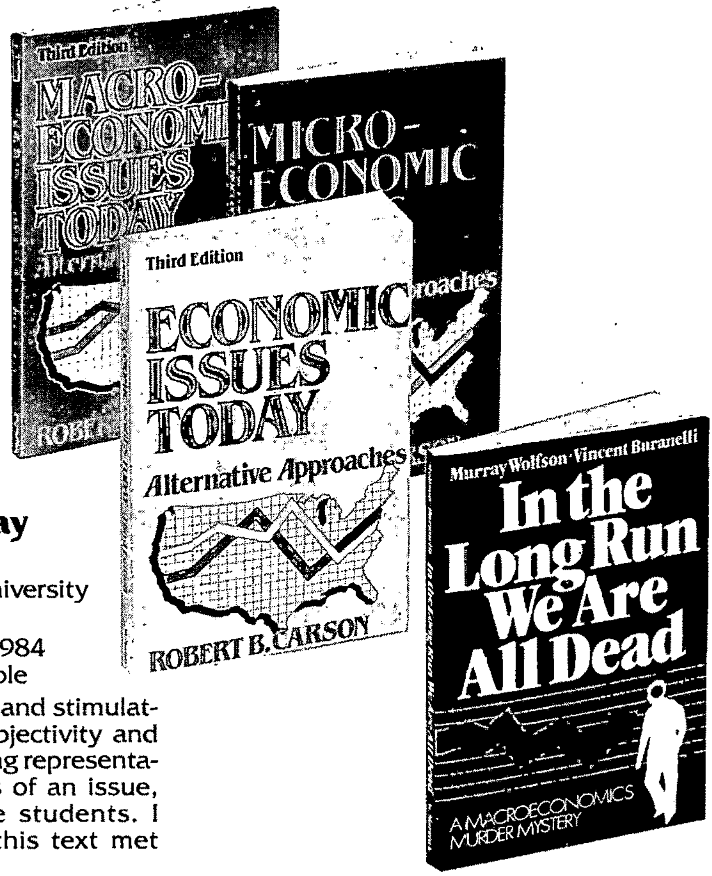
Sylvia Ann Hewlett, Henry Kaufman, and Peter B. Kenen
A United Nations Association of the United States of America book.
Linking effective economic policy to international conditions, this detailed analysis of international economic problems currently confronting the United States examines: international indebtedness, exchange rate regimes, monetary targeting, and trade implications.
October 1984—ca 256 pages—$25.00 (tent.) cloth
ISBN 0-88410-989-5

BALLINGER
PUBLISHING COMPANY
A Subsidiary of *Harper & Row*
in Harvard Square, Cambridge, MA 02138

To order, call toll free 1-800-638-3030

Unusual—and unusually effective





Business Issues Today
Alternative Perspectives

ROBERT B. CARSON, State University College, Oneonta, New York

Paperbound. 348 pages. 1984
Instructor's Manual available

"Very readable, challenging, and stimulating. I appreciate Carson's objectivity and unbiased approach in covering representative viewpoints on all sides of an issue, leaving conclusions to the students. I searched the market and this text met 99% of my expectations!"
—Robin E. Butler, Lakeshore Technical Institute

Economic Issues Today
Alternative Approaches

ROBERT B. CARSON, State University College, Oneonta, New York

Paperbound. 357 pages. 1983
Also available in two separate volumes, *Microeconomic Issues Today* and *Macroeconomic Issues Today*
Instructor's Manual available

"The best text I've seen for giving students a concise, straightforward presentation of the radical, conservative, and liberal views of the major economic issues of the day. Carson is to be congratulated for his balanced approach."
—Michael Wirth, College of Alameda

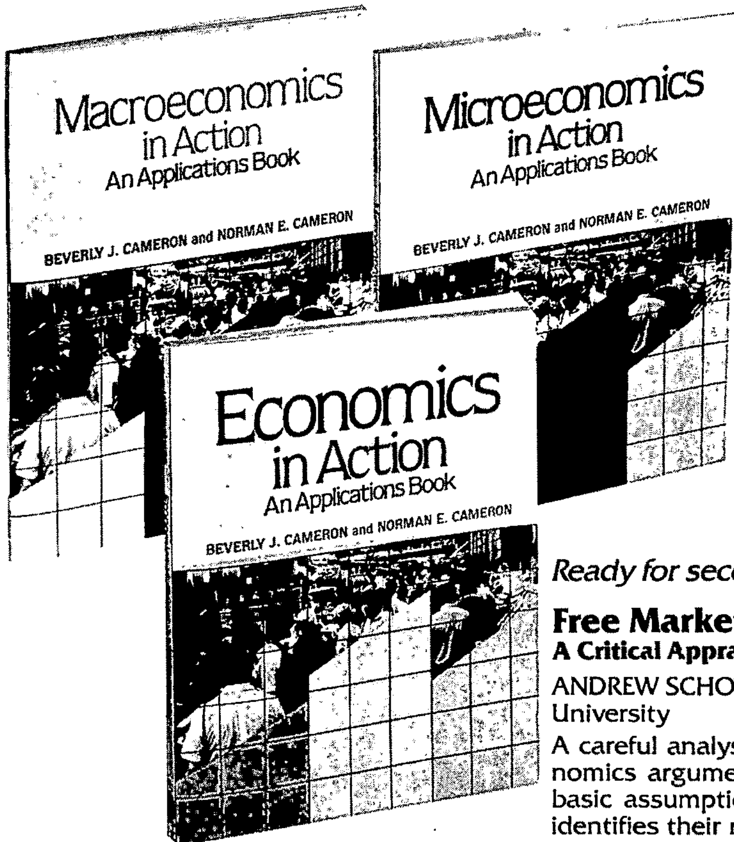
In the Long Run We Are All Dead
A Macroeconomics Murder Mystery

MURRAY WOLFSON, Oregon State University
VINCENT BURANELLI

Paperbound. 177 pages. 1984
Instructor's Manual available

"The authors have done a great job of using a murder mystery to teach macroeconomic theories. The text is well written and easy to understand. Every student of economics should read it. The anticipation of finding out 'who done it' coupled with the skillful way the theories are woven into the story line will hold the student's interest."
—Larry Olarewaju, Virginia State University

—supplements for your courses



Economics in Action
An Applications Book

BEVERLY J. CAMERON and NORMAN E. CAMERON, University of Manitoba

Designed to help students apply economic theory to real-life situations, *Economics in Action* offers a wide selection of newspaper and magazine articles (plus one book excerpt), advertisements, and cartoons dealing with key topics in economics. Questions that relate these source materials to economic concepts encourage students both to recognize theoretical concepts in practice and to analyze real behavior by using those concepts.

Paperbound. 192 pages. 1984
Also available in two separate volumes, *Microeconomics in Action* and *Macroeconomics in Action*
Instructor's Manual available

Ready for second semester—

Free Market Economics
A Critical Appraisal

ANDREW SCHOTTER, New York University

A careful analysis of the free-market economics argument, this book defines the basic assumptions of the argument and identifies their roots in the history of economic thought. Schotter reviews the major criticisms of the argument and raises some serious doubts about the blind advocacy of free market solutions to current social problems.

Paperbound. 148 pages. Publication: September 1984

To request a complimentary examination copy of any of these titles, please write us on your college letterhead specifying your course title, present text, and approximate enrollment. Address your request to:

ST. MARTIN'S PRESS Department JR
175 Fifth Avenue
New York, N.Y. 10010

AMERICAN ECONOMIC ASSOCIATION

1984 ANNUAL MEMBERSHIP RATES

Membership includes:

—a subscription to both *The American Economic Review* (quarterly) plus *Papers and Proceedings* and the *Journal of Economic Literature* (quarterly).

- Regular member with rank of assistant professor or lower or annual incomes of $15,840 or less $33.00
- Regular member with rank of associate professor or annual incomes of $15,840/$26,400 .... $39.60
- Regular member with rank of full professor or annual income above $26,400 .... $46.20
- Junior member (available to registered students for three years only). Student status must be certified by your major professor or school registrar $16.50
- In Countries other than the U.S.A., Add $9.20 to cover postage.
- Family member (second membership without publications; two or more living at same address) ..... $ 6.60

Please begin my issues with:

☐ **March** ☐ **June** (*Includes Papers and Proceedings*) ☐ **September** ☐ **December**

First Name and Initial ______ Last Name ______ Suffix ______

Address Line 1 ______

Address Line 2 ______

City ______

State or Country ______ Zip/Postal Code ______

MAJOR FIELDS (TWO ONLY)
LIST FIELDS WITH WHICH YOU CURRENTLY IDENTIFY. SELECT FIELD CODE FROM *JEL*, "Classification System for Books."

☐ ☐

PLEASE TYPE OR PRINT INFORMATION ABOVE; PLEASE SEND CHECK OR MONEY ORDER PAYABLE IN U.S. DOLLARS. CANADIAN AND FOREIGN PAYMENTS MUST BE IN THE FORM OF A U.S. DOLLAR DRAFT ON A NEW YORK BANK.

Endorsed by (AEA member) ______

Below for Junior Members Only

I certify that the person named above is enrolled as a student at ______

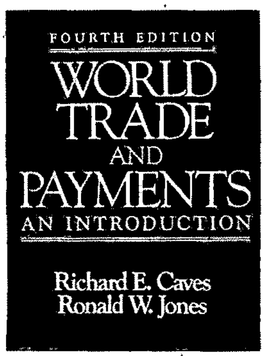
______

Authorized Signature ______

PLEASE SEND WITH PAYMENT TO:

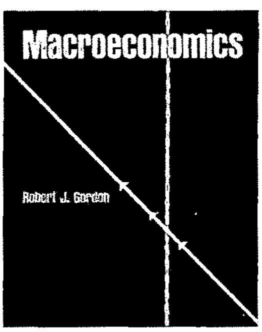
AMERICAN ECONOMIC ASSOCIATION
1313 21ST AVENUE SOUTH, SUITE 809
NASHVILLE, TENNESSEE 37212-2786
U.S.A.

Texts you can count on...



New edition!
WORLD TRADE AND PAYMENTS
An Introduction
Fourth Edition
Richard E. Caves and Ronald W. Jones
Thoroughly updated, this revision of the leading text for international economics features a reworked, up-to-date treatment of trade theory and strengthened coverage of international finance. And this clear, modern account is now more useful than ever — with supportive chapter-ending exercises and problems new to this edition.
Available in November! #132276/cloth/544 pages

New!
MANAGERIAL ECONOMICS
Robert F. Rooney
This comprehensive new text is more than merely a price theory text disguised as a managerial economics text. Rooney's fresh approach presents microeconomics in a more useful way for business students. The text features worked-out cases that show students how theory can be applied to practical business decisions and problems. Calculus is used where appropriate, and numerous problems are included at the end of each chapter.
Available in February! #755966/cloth/480 pages/with Instructor's Manual



Highly successful!
MACROECONOMICS
Third Edition
Robert J. Gordon
Users across the nation agree that Gordon's Third Edition is the best text — with the best package — for the intermediate macro course. Simplified presentation of inflation and unemployment as well as the latest data available enhance this widely-used text's superb integration of modern theory with numerous real-world cases.
#321079/cloth/720 pages/1984/with Instructor's Manual, Student Workbook, and *The Gordon Update* newsletter

MICROECONOMIC THEORY AND APPLICATIONS
Edgar K. Browning and Jacquelene M. Browning
#112232/1983/cloth
604 pages/with Instructor's Manual and Study Guide

PUBLIC SECTOR ECONOMICS
Second Edition
Robin W. Boadway and David E. Wildasin
#100528/1984/cloth
480 pages

ENVIRONMENTAL ECONOMICS AND POLICY
Paul B. Downing
#191809/1984/cloth
350 pages

from Little, Brown.

College Division • 34 Beacon Street • Boston, Massachusetts 02106

Econometric/Statistical Package

xSTAT™

Mainframe Capability, Accuracy and Speed for Your IBM Personal Computer

xSTAT: A Reliable, Economical Alternative to Mainframe Computing

- Is designed for the analysis of **VERY LARGE** as well as small data sets. It has no limit on sample size (total number of observations); and allows up to 50 variables per regression.
- Can generate and save an **X'X product matrix** the first time a data set is used. Subsequent regressions can be made from it to eliminate repetitive readings of input data normally required in statistical packages—very time-consuming when disk inputs are involved.
- Supports Intel 8087 80-bit Math Co-processor for **FAST NUMBER CRUNCHINGS** (has software emulation for systems without the co-processor to attain same accuracy). Double precision is used in matrix operations.
- Produces estimated coefficients that are identical to mainframe TSP results to four significant digits.
- Executes at **HIGH SPEED** when an X'X matrix is used to estimate regressions even for very large samples—it takes only seconds to complete a run.
- The output can be directed to disk, printer or screen. The display can be 132 or 80 columns wide and the statistics and coefficients can be printed in F-format (e.g., 123.456) or E-format (1.23456 + E02).
- Works in either batch or prompt mode; requires no special command language. **EASY TO USE.**

THE PACKAGE INCLUDES:

Data Reformatting and Transformation Program

- Converts formatted data into unformatted data and vice versa; aggregates over records; merges files; accepts data downloaded from mainframe or generated by other programs including Lotus 1-2-3, Visicalc, dBase II, WordStar, other word processors and editors.
- Allows selection of subsamples based on value range or sequential order. Converts fractions to decimal numbers (e.g., stock quotes, 32⅛ => 32.125).
- Has 14 mathematical/logical transformation functions for vector multiplication, division, addition, subtraction, logarithm, exponent, lagged variables, first difference, dummy variables, and time trend.

Descriptive Statistics

- Minimum, maximum, mean, standard deviation, sum of squares, correlation coefficients, **CROSS-TABS**, etc.

Ordinary Least Squares (Multiple regressions) and Two-stage Least Squares

- Generates variance/covariance matrix of coefficients, standard error, t-statistic, R-square, adjusted R-square, Durbin-Watson statistic and F-statistic. Calculates predicted values, residuals, and produces residual plot.
- Allows subsample selection; can also estimate regressions from product moment matrix (contingency table).

Pooling Cross-Section and Time-Series Data

- Estimates regressions for samples with multi-period observations (e.g., panel study). Automatic stacking of data matrix (diagonal block matrix). Estimates coefficients and F-statistics for testing various hypotheses of (1) a common intercept and slope; (2) a common slope and different intercepts; (3) a common intercept and different slopes; (4) different intercepts and different slopes.

SYSTEM REQUIREMENTS

IBM PERSONAL COMPUTER (PC or XT), COMPAQ or IBM PC compatible systems: 1 floppy disk drive or a hard disk system; DOS 1.1 or 2.0 and 192K memory.

PRICE: $250 for the complete package. (Accept checks, Visa & MasterCard. Or order through your local dealer.)

RELATED PRODUCT

Spreadsheet Forecasting Templates for Lotus 1-2-3: reads the estimated coefficients and inverse X'X matrix, saved optionally by xSTAT, directly into 1-2-3 for making "what if" forecasts and producing graphs. No keyboard data entry is needed. The templates are imbedded with formulae for estimating the standard error and confidence interval for the forecasted dependent variable that changes as the forecasted independent variables are varied. **PRICE $100**

MING TELECOMPUTING INC.

Telecommunications and Statistics for Microcomputers

P.O. BOX 101 Lincoln Center, MA 01773 (617) 259-0391

(SAMPLE OUTPUTS ARE AVAILABLE UPON REQUEST.)

Begin making plans to attend the

Ninety-Seventh

Annual Meeting of The American Economic Association

(in Conjunction with Allied Social Science Associations)

to be held in

DALLAS, TX.

Dec. 28-30, 1984

The Employment Center opens Thursday, December 27.

See the Notes section of the September *AER* for the American Economic Association's preliminary program.

The 1985 meeting will be held in New York, NY, December 28-30.

Coming this Fall

Available for your Spring classes

New for 1985!

PUBLIC FINANCE
Harvey S. Rosen,
Princeton University

Revisions for 1985

MANAGERIAL ECONOMICS
Revised Edition
S. Charles Maurice and
Charles W. Smithson, both
of Texas A & M University

THE NEW WORLD OF ECONOMICS
Explorations into the Human
Experience
Fourth Edition
Gordon Tullock, George
Mason University and
Richard B. McKenzie,
Clemson University

PUBLIC POLICIES TOWARD BUSINESS
Seventh Edition
William G. Shepherd, The
University of Michigan

READINGS IN LABOR ECONOMICS AND LABOR RELATIONS
Fifth Edition
Richard L. Rowan, University
of Pennsylvania

15 NOV 1984

Recent Publications

MACROECONOMICS
Third Edition
Paul Wonnacott

MANAGERIAL ECONOMICS
Text, Problems, and Short Cases
Sixth Edition
K. K. Seo

LABOR ECONOMICS
Wages, Employment,
Trade Unionism and Public Policy
Fifth Edition
F. Ray Marshall, Vernon M.
Briggs, Jr., and Allan G. King

Richard D. Irwin, Inc. • Homewood, Illinois 60430

Examination copies for adoption consideration are available upon request; please indicate course title and text presently used.

The American Economic Review

ARTICLES

DECEMBER 1984

THE AMERICAN merican Economic CIATION

•Published at George Banta Co., Inc., Menasha, Wisconsin. The publication number is ISSN 0002-8282.

•*THE AMERICAN ECONOMIC REVIEW* including four quarterly numbers, the *Proceedings* of the annual meetings, the Directory, and Supplements, is published by the American Economic Association and is sent to all members five times a year: March; May; June; September; December.

Regular member dues for 1985, which include a subscription to both the *American Economic Review* and the *Journal of Economic Literature* are as follows:

- $35.00 if annual income is $30,000 or less;
- $42.00 if annual income is above $30,000, but no more than $40,000;
- $49.00 if annual income is above $40,000.

Nonmember subscriptions will be accepted only for both journals: Institutions (libraries, firms, etc.), $100 a year; individuals, $65.00. Single copies of either journal may be purchased from the Secretary's office, Nashville, Tennessee.

In countries other than the United States, add $11.00 to the annual rates above to cover extra postage.

•Correspondence relating to the Directory, advertising, permission to quote, business matters, subscriptions, membership and changes of address should be sent to the Secretary, C. Elton Hinshaw, 1313 21st Avenue So., Suite 809, Nashville, TN 37212-2786. Change of address must reach the Secretary at least six (6) weeks prior to the month of publication. The Association's publications are mailed second class.

•Second-class postage paid at Nashville, Tennessee and at additional mailing offices, Printed in U.S.A.

•Postmaster: Send address changes to *American Economic Review*, 1313 21st Avenue So., Suite 809, Nashville, TN 37212-2786.

Founded in 1885

Officers

President
CHARLES L. SCHULTZE
The Brookings Institution

President-Elect
CHARLES P. KINDLEBERGER
Massachusetts Institute of Technology

Vice Presidents
ZVI GRILICHES
Harvard University
ROBERT L. HEILBRONER
New School for Social Research

Secretary
C. ELTON HINSHAW
Vanderbilt University

Treasurer
RENDIGS FELS
Vanderbilt University

Managing Editor of The American Economic Review
ROBERT W. CLOWER
University of California-Los Angeles

Managing Editor of The Journal of Economic Literature
MOSES ABRAMOVITZ
Stanford University

Executive Committee

Elected Members of the Executive Committee
ANN F. FRIEDLAENDER
Massachusetts Institute of Technology
JOSEPH E. STIGLITZ
Princeton University
WILLIAM D. NORDHAUS
Yale University
A. MICHAEL SPENCE
Harvard University
VICTOR R. FUCHS
Stanford University
JANET L. NORWOOD
Bureau of Labor Statistics

EX OFFICIO Members
GARDNER ACKLEY
The University of Michigan
W. ARTHUR LEWIS
Princeton University

391.001

THE AMERICAN ECONOMIC REVIEW

ROBERT W. CLOWER
Managing Editor

JOHN G. RILEY
Associate Editor

WILMA ST. JOHN
Production Editor

THERESA DE MARIA
Assistant Editor

Board of Editors

GEORGE A. AKERLOF
CLIVE D. BULL
PATRICIA DANZON
MICHAEL R. DARBY
PHILIP E. GRAVES
JACK HIRSHLEIFER
MEIR KOHN
FREDERIC S. MISHKIN
SHERWIN ROSEN
RICHARD SCHMALENSEE
SUSAN WOODWARD
LESLIE YOUNG

• Manuscripts and editorial correspondence relating to the regular quarterly issues of this *Review* and the *Papers and Proceedings* should be addressed to Robert W. Clower, Managing Editor, *AER* Editorial Office, University of California, Los Angeles, CA 90024. Manuscripts should be submitted in triplicate and in acceptable form, and should be no longer than 50 pages of double-spaced typescript. A submission fee must accompany each manuscript: $25 for members; $50 for nonmembers. *Style Instructions* for guidance in preparing manuscripts will be provided upon request to the editor.

• No responsibility for the views expressed by authors in this *Review* is assumed by the editors or the publishers, The American Economic Association.

• Copyright © American Economic Association 1984. All rights reserved.

DECEMBER 1984

VOLUME 74, NUMBER 5

Articles

Shorter Papers

The following Statement of Ownership, Management, and Circulation is provided in accordance with the requirements, as contained in 39 U.S. Code 3685. The *American Economic Review* is owned, managed, and published by the American Economic Association, a nonprofit educational organization, located at 1313 21st Avenue So., Nashville, Davidson County, Tennessee 37212. The Managing Editor is Professor Robert W. Clower, *American Economic Review*, Bunche Hall, 405 Hilgard, University of California, Los Angeles, CA 90024. During the preceding 12 months the average number of copies printed for each issue was 27,700, the average paid circulation, 23,806; the average free distribution, 390; the average number of copies distributed, 24,196. Corresponding figures for the last issue before filing: 27,500 total number copies printed; 24,400 total paid circulation; 385 free copies distributed; 24,785 total distribution.

Orange Juice and Weather

By RICHARD ROLL*

Frozen concentrated orange juice is an unusual commodity. It is concentrated not only hydrologically, but also geographically; more than 98 percent of U.S. production takes place in the central Florida region around Orlando.[1] Weather is a major influence on orange juice production and unlike commodities such as corn and oats, which are produced over wide geographical areas, orange juice output is influenced primarily by the weather at a single location. This suggests that frozen concentrated orange juice is a relatively good candidate for a study of the interaction between prices and a truly exogenous determinant of value, the weather.

The relevant weather for OJ production is easy to measure. It is reported accurately and consistently by a well-organized federal agency, the National Weather Service of the Department of Commerce. Forecasts of weather are provided by the same agency and this makes it possible to assess the predictive ability of OJ futures prices against a rather exacting standard.

Geographic concentration is the most important attribute of orange juice for our empirical purposes, but the commodity also possesses other convenient features. It seems unlikely to be sensitive to *non*weather influences on supply and demand. For example, although the commodity is frozen and not very perishable, only a small amount is carried over in inventory from one year to the next. During 1978, for example, inventory declined to about 20 percent of the year's "pack" of new juice.[2]

Data on short-term variability in demand are nonexistent, but there is little reason to suspect much. Orange juice demand might very well respond to price variation in substitutes such as, say, apple juice; but national income and tastes probably do not fluctuate enough to explain a significant part of the *daily* OJ juice movement[3] (which is substantial, as we shall see).

Short-term variations in supply induced by planting decision must also be quite low because of the nature of the product. Oranges grow on trees that require five to fifteen years

*Graduate School of Management, University of California, Los Angeles, CA 90024. I am grateful for discussions with Eugene Fama and Stephen Ross, for comments on an earlier draft by Gordon Alexander, Thomas Copeland, Michael Darby, David Mayers, Huston McCulloch, and Sheridan Titman, for the cooperation of Paul Polger of the National Oceanographic and Atmospheric Administration, and for comments in seminars from the finance faculties of the universities of British Columbia, Alberta, and Illinois. Kathy Gillies provided excellent research assistance. Financial assistance was provided by Allstate, the Center for Research in Financial Markets and Institutions at UCLA and by the Center for the Study of Futures Markets at Columbia.

[1]The proportion produced in Florida is now close to 100 percent. Indeed, the annual publication, *Agricultural Statistics*, by the U.S. Department of Agriculture, no longer gives a breakdown by area, reporting the production only for Florida (presumably because production elsewhere is so small). The last breakdown by area was for 1961 (see *Agricultural Statistics*, 1972, Table 324). In 1961, Florida produced 115,866,000 gallons while California and Arizona combined produced 2,369,000 gallons. It may surprise the reader to know that OJ production for frozen concentrate is mainly a Florida industry; many *table* oranges do come from California. This difference between Florida and California oranges is attributable to differences in their sugar and juice content and in their exteriors. Florida oranges are sweeter and make better-tasting juice. California oranges, being less sweet, have a longer shelf life and they also tend to have less juice but more appealing skins. Apparently, there is not as much substitutability as might have been imagined. Actually, Florida produces the bulk of all oranges for both table and juice. In 1972–73, for example, Florida orange production by weight was about 80 percent of the U.S. total. (See *Florida Agricultural Statistics*, Table 3, p. 4.)

[2]See Tables 380 and 382 of *Agricultural Statistics* (1979, pp. 252 and 254).

[3]A rough indication of exogenous shifts in demand due to income and tastes can be obtained from U.S. consumption of all citrus fruit which has hovered closely around 27 pounds per capita for a number of years (see Table 384, p. 255, *Agricultural Statistics*, 1979).

to mature.[4] Thus, any vagaries in farmers' planting decisions are felt much later and do not impact the current year's crop. There might, however, be short-term effects from farming decisions concerning fertilizer use or harvesting methods. These could be influenced by the prices of fertilizer and energy.

It should be emphasized that even unstable conditions of demand and supply would not eliminate the influence of weather, they would simply make that influence harder to measure empirically. The main argument in favor of studying orange juice instead of other commodities is the geographical concentration of OJ production. The fact that nonweather influences seem unlikely to generate much empirical noise is simply an added benefit.

I. Data

A. *Orange Juice Futures*

Futures contracts in frozen concentrated orange juice are traded by the Citrus Associates of the New York Cotton Exchange. There are usually nine contracts outstanding with deliveries (expirations) scheduled every second month, January, March, etc., the most distant delivery being 17 to 18 months from the present. A contract is for 15,000 pounds of orange solids standardized by concentration (termed "degrees Brix") and with minimum "scores" for color, flavor, and defects.[5]

Price data[6] are available for each day since the exchange began OJ trading in the early 1970's. However, the weather data are available only for October 1975 through December 1981, so this constitutes the sample period. There were 1,564 trading days during this period.

As is typical of many commodities, trading volume in OJ futures tends to be concentrated in the near-maturity contracts. The open interest of distant contracts, say 8 to 18 months maturity, is often only 10 percent or less of the open interest in nearer contracts, say from 2 to 6 months maturity. Because of well-known problems in price data from thin markets,[7] the fourth and longer maturities were discarded in the following empirical work.

The nearest-maturity contract was also discarded after a close examination of its price behavior around the maturity date. Volume of trading is quite high in the nearest contract until just a few days before expiration. But in the last several days of the contract's life, open interest declines and price volatility increases substantially. A good example of the ensuing econometric problem involved the contract which matured on November 16, 1977. During the last fifteen minutes before expiration, its price rose from \$1.30 to \$2.20 per pound, an annualized rate of return of about 1.8 *million* percent. Such events would seem to have little to do with the weather.

This leaves us with two contracts having, respectively, between 2 and 4 months and between 4 and 6 months to maturity; an equally weighted average of the daily returns on these two contracts was chosen as the basic OJ return for use in all subsequent analysis. (Using either contract separately gives virtually identical results. This is to be expected because the correlation between their returns is .97.)

On a contract expiration day, the shorter of these two contracts is dropped and a new contract, previously the fourth-from-

[4]See John McPhee (1967) for a fascinating and entertaining description of orange tree propagation and of the citrus business in general.

[5]The contract quality is specified as follows: "U.S. Grade A with a Brix value of not less than 51° having a Brix value to acid ratio of not less than 13 to 1 nor more than 19.0 to 1 and a minimum score of 94, with the factor of color and flavor each scoring 37 points or higher, and defects at 19 or better..., provided that [OJ] with a Brix value of more than 66° shall be calculated as having 7.278 pounds of solids per gallon" (*Citrus Futures*, undated). "Degrees Brix" is a term used in honor of a nineteenth-century German scientist, Adolf F. W. Brix (McPhee, p. 129).

[6]The price used here is the "settlement" price. This price (which may or may not reflect an actual transaction) is determined by members of the exchange at the close of each day's trading. It is the price reported in the financial press.

[7]See Myron Scholes and Joseph Williams (1977), Elroy Dimson (1979), Marshall Blume and Robert Stambaugh (1983).

TABLE 1—OJ FUTURES DAILY RETURNS BY DAY OF WEEK AND BY SEASON OCTOBER 1975–DECEMBER 1981

Day of Week	Mean Returns[a]				
	Winter[b]	Spring	Summer	Autumn	All Seasons
Monday[c]	−.256	−.321	−.107	.0309	−.158
	(2.58)	(1.84)	(1.52)	(1.84)	(1.96)
Tuesday	.224	.269	.199	−.107	.146
	(2.11)	(1.37)	(.147)	(1.48)	(1.62)
Wednesday	.301	.188	−.102	−.169	.0540
	(1.72)	(1.54)	(1.40)	(1.36)	(1.52)
Thursday	.167	−.219	.113	.153	.0518
	(2.14)	(1.16)	(1.21)	(1.35)	(1.51)
Friday	.290	.0227	−.125	.242	.108
	(1.98)	(1.55)	(1.63)	(1.53)	(1.68)
Post-Holiday	−.0554	.311	.278	−.0817	.0102
	(1.78)	(1.72)	(1.25)	(1.37)	(1.52)
All Days	.141	−.00741	−.00079	.0253	.0392
	(2.09)	(1.51)	(1.51)	(1.52)	(1.66)

Notes: Levene's test (see Morton Brown and Alan Forsythe, 1974) for equal variances: $F = 3.59$; tail probability $\cong 0$. Dummy variable regression:

$$\underset{(1.86)}{R_t = .0886} - \underset{(-2.30)}{.247}\ d_m - \underset{(.328)}{.0784}\ d_h \qquad R^2 = .00211$$

where d_m is 1 on Monday, 0 otherwise, and d_h is 1 on post-holiday day, zero otherwise.

[a]Average of the second- and third-nearest maturity contracts' returns. The mean returns (standard deviations) of the two contracts separately were .0388(1.70) and .0397(1.65), respectively; their correlation was .969. The returns are shown in percent; standard deviations are shown below in parentheses.

[b]Winter is defined as December, January, February, inclusive. Spring, Summer, and Autumn include, respectively, each subsequent three months.

[c]Monday returns are from settlement price Friday to settlement price Monday. Other days are from settlement on previous day. Post-Holiday returns are from settlement on day before holiday to close on day after holiday.

the shortest maturity, starts to be used in construction of the return series. The return on the new contract over the expiration date replaces the return on what has become the shortest maturity contract.[8]

Table 1 gives information about OJ returns over the sample period. The grand mean return is .0392 percent per day, about 10.3 percent per annum. The rather large volatility of these returns is shown by the fact that the standard error of the mean daily return is $1.66/(1563)^{1/2} = .0420$. The standard error is larger than the mean despite the large sample size.

In the body of the table, means and standard deviations are broken down by season and by day of the week. The seasonal pattern shows a larger mean and larger variability during winter. This might have been anticipated on the grounds that colder temperatures and the risk of freezing make investments in orange juice more hazardous during the winter months. A finer breakdown indicates, however, that the larger winter mean OJ return is due to January alone, perhaps for the same reason that equities of small

[8]Specifically, let $R_{T,t}$ be the continuously compounded return on day t of a contract which matures on calendar date T. Say that contracts mature on days $T = 60, 120, 180, 240, 360$. The return series ($R_t^*$) used here is calculated as follows

$$R_t^* = (R_{120,t} + R_{180,t})/2 \qquad t \leq 60$$

$$R_t^* = (R_{180,t} + R_{240,t})/2 \qquad 120 \geq t > 60$$

$$R_t^* = (R_{240,t} + R_{360,t})/2 \qquad 180 \geq t > 120,$$

and similarly as times goes on and contracts mature.

firms have larger January returns.[9] (Compare Donald Keim, 1983.)

The day-of-the-week results can be compared to recent work on equity returns (Kenneth French, 1980; Michael Gibbons and Patrick Hess, 1981) which found a significantly negative Monday effect. A similar pattern is observed here in the means.[10] Thus, insofar as mean returns are concerned, OJ futures seem to display annual and weekly seasonals similar to equities.

The intraweek pattern of standard deviations is interesting for what it does *not* display. Since Monday's return covers a three-day period, while other days of the week cover only 24 hours, one might have thought that Monday's variance of returns would be approximately three times as large as the other days. Yet the ratio of Monday's to the average of the other days' variances is only about $(1.96/1.58)^2 = 1.54$. Monday's return has too low a variance. (Note that post-holiday returns, which are always for at least two calendar days, also have too low a variance.) Because of this pattern of variances across days, it must be admitted that weather may not be the only relevant factor for OJ returns after all. If weather alone were moving OJ prices, Monday's return volatility should be larger because weather surprises must occur just as readily on a weekend as on any other day. Nevertheless, since no one has yet discovered just what factors *are* causing day-of-the-week patterns, I shall proceed with an examination of weather, which is at least a known factor.

The OJ futures exchange imposes limits on price movements. These limit rules (see Table 2) prevent the price from moving by more than a certain amount from the previous day's settlement price. When a significant event, such as a freeze in Florida, causes the price to move the limit, the settlement price on that day cannot fully reflect all available information. In other words, limit rules cause a type of market information inefficiency (but not a profit opportunity). This might be inconsequential if limit moves occurred rarely; unfortunately, they are rather common. During the October 1975–December 1981 period, one or both of the two contracts being used here moved the price limit on 160 different trading days, slightly over 10 percent of the trading days in the sample. This implies that about 10 percent of the recorded prices in the sample are known in advance not to reflect all relevant available information.

Limit rules might be suspected as the reason why Monday's variance is too low since these rules would be more frequently applied to limit the three-day weekend/Monday return. It turns out, however, that only 40 of the 160 limit moves in the sample occurred on Monday. This frequency is slightly higher than the frequency of 20 percent which would be expected if all five weekday returns

TABLE 2—LIMIT RULES OF THE CITRUS ASSOCIATES OF THE NEW YORK COTTON EXCHANGE AFTER (BEFORE) JANUARY 1, 1979

General Rule: Prices may move no more than 5 (3) cents per pound, \$750 (\$450) per contract, above or below the settlement price of the previous market session.

Increased Limit Rule: When three or more contract months have closed at the limit in the same direction for three successive business days, the limit is raised to 8 (5) cents per pound for those contract months. The limit remains at 8 (5) cents until fewer than three contract months close at the limit in the same direction, then the limit reverts to 5 (3) cents on the next business day.

Current Rule for Near Contract: On the last three days before the near contract's expiration, its limit is 10 cents per pound. If that limit is reached during the market session, trading is suspended on *all* contracts for fifteen minutes. Then another 10 cents is added to or deducted from the near contract's limit and trading recommences. Limit moves and fifteen-minute suspensions can be repeated until the market's close. If this happens on the last day before expiration, trading hours are extended.

[9] January's average daily OJ return was .701 percent (standard error = .238) while all other months combined had an average daily return of −.0193 percent (standard error = .0402).

[10] When compared against other days of the week in an analysis of variance, Monday's return is found to be significantly lower (*F*-statistic of 5.20 and tail probability of .0228). Monday's mean return is, however, only marginally significantly negative; the standard error of the mean (of −.158) is .114 percent. The dummy variable regression reported at the bottom of the table shows that the Monday effect is significant but that the explained variance is low.

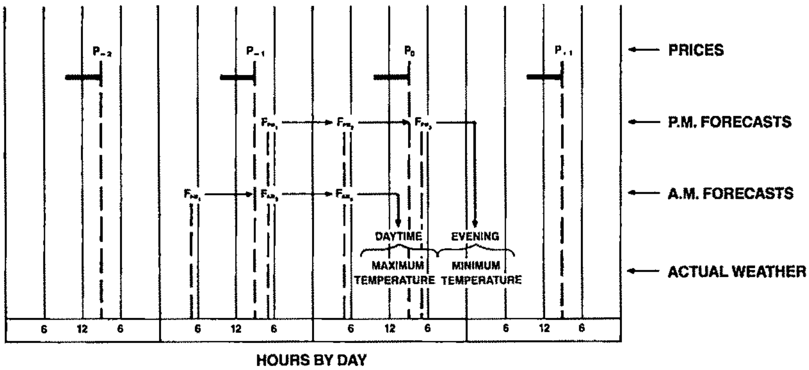


FIGURE 1. TIMING SCHEMATIC OF OJ FUTURES MARKET, WEATHER FORECASTS, AND ACTUAL PERIOD OF WEATHER AT ORLANDO

Note: ▬ indicates market trading hours

covered the same number of hours. The ratio of Monday's return variance to the average variance on the other days is only 1.75 even when all limit move observations are excluded.

B. *Central Florida Weather*

The U.S. Weather Service reporting station in Orlando issues a variety of different weather bulletins. The most relevant information for oranges involve temperature and rainfall; the data[11] used here consist of daily information on these two variables.

Each 24-hour interval is divided into 12-hour daytime and evening periods. The daytime period begins at 7:00 A.M., eastern standard time, and ends at 7:00 P.M. on the same day. The evening period begins at 7:00 P.M. and ends at 7:00 A.M. the following day. For the daytime period, the weather service reports actual rainfall and the *maximum* temperature, while for the evening period, the rainfall and *minimum* temperature are reported.

Three different forecasts of both rainfall and temperature are also provided. They correspond to periods 36 hours, 24 hours, and 12 hours in advance of the 12-hour period to which the forecast applies. For example, say that the forecast is of the maximum temperature on January 5 (which could occur anytime from 7:00 A.M. until 7:00 P.M.). The first forecast is issued about 5:00 A.M. on January 4. (I call this the 36-hour-ahead forecast because it is developed and issued during the third 12-hour period prior to the 12-hour observation period of the actual maximum temperature.) A second forecast applying to the maximum January 5 temperature is issued at 5:00 P.M. on January 4; then, the third forecast is issued at 5:00 A.M. on January 5. This same cycle, but delayed by 12 hours, is used to issue forecasts of the minimum temperature on January 5 (from 7:00 P.M. January 5 until 7:00 A.M. January 6). Rainfall forecasts for the daytime and evening periods are issued along with the temperature forecasts.

Figure 1 gives a timing schematic of the actual weather, the forecasts of weather, and the trading times of orange juice futures. The symbol p_0 indicates the OJ settlement price on a particular calendar date. Note that p_0 is observed during the 12-hour daytime period, well before the evening period begins, and even before the last forecast of evening weather issued by the weather service. For this reason, we might anticipate that surprises in daytime weather would be

[11]The cooperation of Paul Polger of the National Oceanographic and Atmospheric Administration, who provided these data and provided a detailed explanation, is gratefully acknowledged.

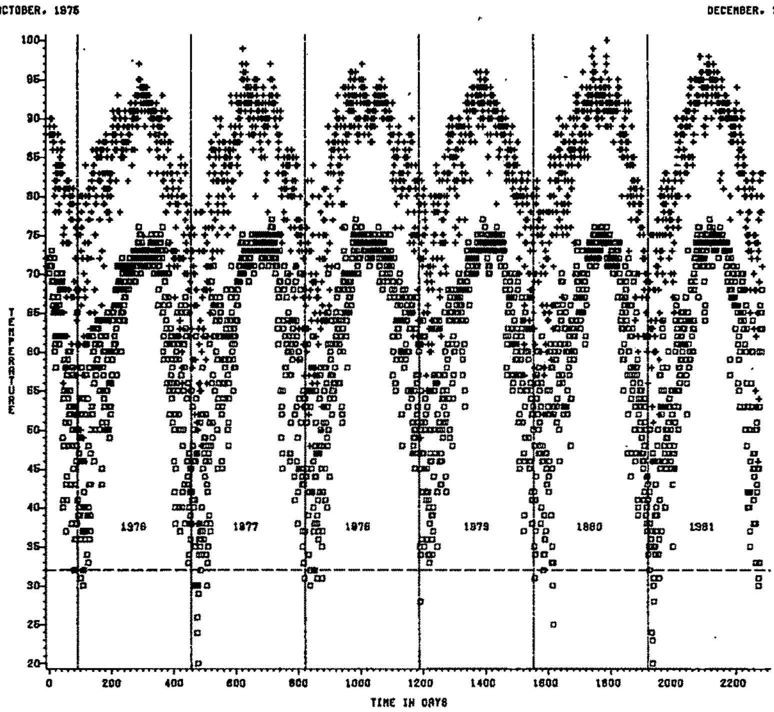


FIGURE 2. MAXIMUM AND MINIMUM DAILY TEMPERATURES AT ORLANDO

associated with price movements of p_{-1} to p_0 while evening weather surprises would influence price changes p_0 to p_{+1}.

The actual daily temperatures are plotted for the sample period in Figure 2 (+ indicates daily maximum and □ indicates minimum). The figure shows that temperatures in central Florida are not only lower during the winter season, they are also more variable. Damage to orange trees occurs if the temperature drops below freezing and stays there for a period of several hours. Thus, the minimum (P.M.) temperature during the winter months would seem to be an important factor influencing the size of the crop and the price of futures.

Table 3 shows that the Weather Service's short-term forecasts of temperature are quite accurate on average and that the forecast improves as its period approaches.[12] The OJ futures market has access to both the 36-hour-ahead and the 24-hour-ahead forecast of that day's P.M. minimum temperature (compare Figure 1). These two forecasts are issued prior to the market's opening. Thus, even aside from whatever private weather forecasts are made by OJ futures traders, two reasonably accurate forecasts of the day's

[12] However, there is a curiosity in these forecasts. Note that the A.M. level regressions tend to have slopes ($\hat{b}$) below 1.0. This could be due to errors in the data (rather than in the forecasts). The data were filtered and obvious transcription errors were corrected as detected. Of course, there may still be errors remaining. Errors-in-variables-induced attenuation bias cannot, however, explain why the P.M. forecast intercept is significantly negative. The Theil inequality proportions indicate significant bias in the P.M. forecasts. Note that the low Durbin-Watson statistics on the 36-hour-ahead forecasts are to be expected since there is an intervening actual between this forecast and the actual to which the forecast applies. (See Figure 1.) In other words, the 36-hour-ahead forecast on day t is issued before the forecast error is known for the 36-hour-ahead forecast from day $t-1$. This induces positive dependence in adjacent forecast errors.

TABLE 3—TEMPERATURE FORECAST ACCURACY FOR ORLANDO
OCTOBER 1975–DECEMBER 1981

Hours Forecast is Ahead[a]	Temperature Level $\hat{a}$ (1)	$\hat{b}$ (2)	R^2 (3)	U^m (4)	Temperature Change $\hat{a}$ (1)	$\hat{b}$ (2)	R^2 (3)	U^m (4)
Maximum (A.M.) Temperature Forecast								
36	4.23	.953	.872	1.15	.357	.832	.604	.777
(2,040)	(6.34)	(118.)	(1.53)	(1.60, 97.3)	(3.34)	(55.7)	(1.81)	(5.82, 93.4)
24	4.60	.951	.896	3.24	.667	.912	.663	2.57
(2,049)	(7.79)	(133.)	(1.81)	(2.16, 94.6)	(6.73)	(63.4)	(1.97)	(1.75, 95.7)
12	4.32	.952	.911	1.90	.511	.984	.708	1.55
(2,048)	(7.96)	(145.)	(1.91)	(2.46, 95.6)	(5.56)	(70.3)	(1.90)	(.061, 98.4)
Minimum (P.M.) Temperature Forecast								
36	−1.48	1.01	.884	6.14	−1.62	.771	.495	6.28
(2,048)	(−2.93)	(125.)	(1.42)	(.035, 93.8)	(−9.24)	(44.8)	(1.64)	(7.49, 86.2)
24	−2.71	1.03	.907	8.88	−1.89	.823	.575	8.86
(2,038)	(−5.92)	(141.)	(1.58)	(.532, 90.6)	(−11.7)	(52.5)	(1.64)	(5.35, 85.8)
12	−.852	1.00	.922	6.23	−1.49	.902	.648	5.92
(2,048)	(−2.11)	(155.)	(1.76)	(0.0, 93.8)	(−10.2)	(61.3)	(1.82)	(2.01, 92.1)

Notes: Regression: Actual $= \hat{a} + \hat{b}$ (forecast). The "actual" is the minimum or maximum temperature observed during a 12-hour period. In the "changes" regression, the dependent variable is the actual percentage change from the previous day's corresponding 12-hour period and the explanatory variable is the predicted percentage change.

Cols. (1),(2): t-statistics are shown in parentheses.

Cols. (3): Durbin-Watson statistics are shown in parentheses.

Cols. (4): U^r, U^d are shown in parentheses. The inequality proportions are shown in percent. See Henri Theil (1966, pp. 32–34). U^m = bias, U^r = regression, U^d = disturbance, proportions of mean squared prediction error due to, respectively, bias, deviation of regression slope from 1.0, and residuals.

[a]Sample size is shown in parentheses. There were 2,284 calendar days in the sample. However, the data contain numerous missing observations.

crucial minimum temperature are publicly available during trading hours.

Rainfall is also predicted by the Weather Service, but the form of the forecast is less useful for our purposes than in the case of temperature. The forecast "probability" of rain is always an even decile such as 30 percent and it rarely exceeds 60 percent. Weather service officials have told me that this forecast is intended to convey the chance of *any* measurable precipitation.

Table 4 reports the complete sample distribution of rainfall forecasts and actuals (the latter are provided in categories only). As shown, high forecast probabilities of rain are unusual even though there is measurable rainfall during about 28 percent of the reporting periods. The last column shows that the actual frequency of the rain is not far from the forecast probability. There is not a strong connection between the forecast probability and the *amount* of rain, but the Weather Service forecast is not intended to predict the amount, simply the chance of rain in *any* amount.

As shall be shown in the next section, there is an obvious relation between temperature and the price of OJ futures. The relation between rainfall and price is much more difficult to detect, if it is there at all. Perhaps this is due to temperature being a more important variable for the crop. Perhaps it is due to less useful weather data regarding rainfall.

TABLE 4—FORECAST PROBABILITY OF RAIN VS. ACTUAL RAINFALL BY CATEGORY IN ORLANDO OCTOBER 1975–DECEMBER 1981

Forecast Probability of Rain[a]	Actual Rainfall (inches) 0	.001–.009	.01–.120	.121–.25	.251–.50	.501–1.0	1.01–2.0	2.01–3.0	3.01–4.0	Total	Frequency of Measurable Precipitation[a]
	Frequency (All Forecasts)										
0	3157	79	28	12	3	1	1	0	0	3281	3.78
10	2439	216	100	39	29	14	9	1	0	2847	16.7
20	1401	266	153	51	34	17	11	2	0	1935	27.6
30	904	260	180	83	39	34	17	2	0	1519	40.5
40	420	178	156	68	58	56	35	7	1	979	57.1
50	279	133	177	80	72	59	40	6	5	851	67.2
60	116	70	120	63	48	37	30	0	3	487	76.2
70	18	22	29	22	16	23	7	1	0	138	87.0
80	8	4	6	2	5	12	3	1	0	41	80.5
90	1	1	7	3	0	5	1	0	0	18	94.4
100	0	0	1	0	0	1	1	0	0	3	100.
Total	8743	1229	957	423	304	259	155	20	9	12099	

Note: χ^2 Test of Dependence:

	χ^2	Tail Probability	Forecasts	χ^2	Tail Probability
All Forecasts	4151	$p=0.0$	36-Hours-Ahead	1185	$p=0.0$
All A.M. Forecasts	2277	$p=0.0$	24-Hours-Ahead	1421	$p=0.0$
All P.M. Forecasts	1559	$p=0.0$	12-Hours-Ahead	1744	$p=0.0$

[a]Shown in percent.

II. Empirical Results

A. *Temperature*

Cold weather is bad for orange production. Orange trees cannot withstand freezing temperatures that last for more than a few hours. Florida occasionally has freezing weather and the history of citrus production in the state has been marked by famous freezes. In 1895, almost every orange tree in Florida was killed to the ground on February 8, production declined by 97 percent, and 16 years passed before it recovered to its previous level.[13] Farmers have since learned how to counter freezes with hardier trees, smudge pots, water spraying,[14] and air circulation by large fans; but although the trees are now more likely to survive a freeze, the crop can be severely damaged. Even a mild freeze will prompt the trees to drop significant amounts of fruit.

Figure 3 illustrates the impact of freezing weather on OJ futures prices during the sample period. The actual minimum temperature at Orlando is plotted along with the OJ price level.[15] Freezing level is indicated by the horizontal dashed line.

During this $6\frac{1}{4}$-year period, there were 27 recorded freezing temperatures (below 32°) at Orlando out of 2284 calendar days. However, only four periods registered temperatures below 30°. These occurred on January 17–21, 1977, January 2, 1979, March 2, 1980, January 11–13 and 18, 1981. (See Figure 2 also.) Figure 3 shows that these episodes were accompanied by significant price increases. The January freezes in 1977 and 1981 were particularly harsh in that six successive days and three successive nights,

[13]McPhee (p. 101).

[14]Spraying trees with water during a freeze can protect them under certain conditions. The water, freezing on the trees' leaves and buds, gives off heat in the process of changing from a liquid to a solid.

[15]Thirty cents has been added to the OJ price in order to keep the plots apart. The price is an average of the second and third shortest-maturity contracts. (See Section I.)

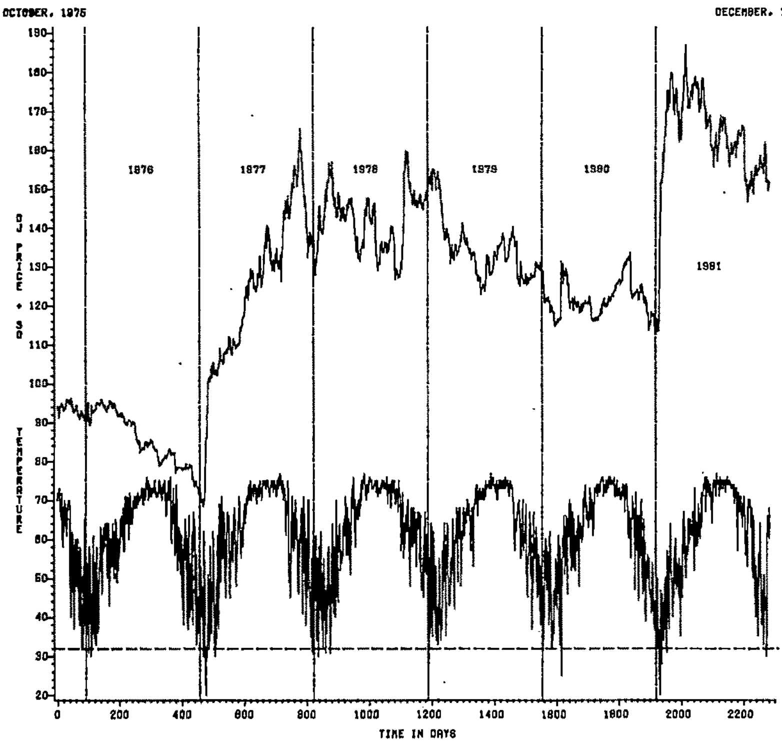


FIGURE 3. OJ FUTURES PRICES AND MINIMUM TEMPERATURES AT ORLANDO

respectively, had freezing temperatures. The most severe freeze during this sample, and the largest accompanying price increase, occurred during the latter period, on January 11, 12, and 13, 1981, when successive daily minimum temperatures were 24°, 23°, and 20°. During the week of January 12–16, OJ futures prices were up the limit on all five trading days.

Market participants realize, of course, that severe freezes are more likely during winter, so the price of OJ futures in the autumn should be high enough to reflect the probability of a freeze during the coming season. Each day thereafter that passes without a freeze should be accompanied by a slight price decline, a relief that winter is one day closer to being over. Also, harvesting of oranges begins in the fall and lasts until early summer, and inventories typically increase over the winter months.

For both of these reasons: freezes that do not occur and inventory build-up; there is a downtrend in futures prices during a typical nonfreeze winter. This pattern can be seen in every year of the sample (Figure 3), except 1977. A general downward movement with small fluctuations is interrupted by occasional sharp price increases sufficient to bring positive returns, on average, to those with long positions.[16] The distribution of returns is very skewed to the right.

If the OJ futures market is an efficient information processor, it should incorporate all publicly available long-term and short-term weather forecasts. Any private forecasts should be incorporated to the extent that traders who are aware of those forecasts are also in command of significant resources. The futures price should, therefore, incorporate the predictable part of weather in advance. Unpredicted weather alone should be

[16]An extensive theoretical discussion of this phenomenon is given by Benoit Mandelbrot (1966).

TABLE 5—OJ FUTURES RETURNS AND TEMPERATURE FORECAST ERRORS WITH AND WITHOUT WEIGHTING, OCTOBER 1975–DECEMBER 1981

Seasons	Hours Forecast is Ahead[a]	Return Lead/Lag (Days) b_{-2}	b_{-1}	b_0	b_{+1}	b_{+2}
Maximum (A.M.) Temperature Forecast						
Unweighted	36	.105	−.00414	−.0463	−.00397	−.322
	(1,391)	(1.31)	(−.0507)	(−.567)	(−.0487)	(−3.91)
Weighted		.102	−.0558	−.0894	−.0600	−.490
		(1.15)	(−.624)	(−1.00)	(−.673)	(−5.37)
Unweighted	24	.0639	−.0497	−.0113	.0379	−.247
	(1,408)	(.872)	(−.673)	(−.154)	(.510)	(−3.36)
Weighted		.0374	−.0615	.0224	.0585	−.379
		(.461)	(−.750)	(.275)	(.714)	(−4.71)
Unweighted	12	.000123	−.0715	−.000467	.0565	−.123
	(1,400)	(.00186)	(−1.07)	(−.00699)	(.838)	(−1.84)
Weighted		−.0851	−.0905	.00691	.0295	−.191
		(−1.17)	(−1.23)	(.0936)	(.398)	(−2.62)
Minimum (P.M.) Temperature Forecast						
Unweighted	36	.0822	−.104	−.154	.136	.0570
	(1,407)	(.632)	(−.791)	(−1.17)	(1.03)	.436
Weighted		.101	−.198	−.379	.133	.0561
		(.664)	(−1.30)	(−2.49)	(.874)	(.374)
Unweighted	24	.0412	−.139	−.352	−.238	−.0404
	(1,399)	(.357)	(−1.20)	(−3.03)	(−2.03)	(−.348)
Weighted		.0593	−.220	−.673	−.544	−.139
		(.442)	(−1.62)	(−4.96)	(−3.99)	(−1.09)
Unweighted	12	−.0698	−.152	−.263	−.0849	.104
	(1,398)	(−.677)	(−1.47)	(−2.52)	(−.807)	(.991)
Weighted		−.0796	−.231	−.549	−.217	.133
		(−.678)	(−1.97)	(−4.62)	(−1.83)	(1.13)

Notes: The regression equation is $\log(A/F)_t = a + b_{-2}R_{t-2} + b_{-1}R_{t-1} + b_0 R_t + b_1 R_{t+1} + b_2 R_{t+2}$, where A is actual temperature, F is forecast temperature and R_t is the return on day t of an equally weighted sum of two futures contracts.

T-statistics are shown in parentheses. All Durbin-Watson statistics were in the range 1.6–1.9. Adjusted R^2s were between 1 and 3 percent.

The weighting scheme is January = 7, February = 6, March = 5, April = 4, May = 3, June = 2, July = 1, August = 2, September = 3, October = 4, November = 5, December = 6.

[a]Sample size is shown in parentheses.

contemporaneously correlated with price movements.

To examine the market's information processing ability, a series of empirical tests were carried out relating surprises in temperature to OJ futures price changes. The temperature forecast error, the percentage difference between the actual temperature and the forecast temperature provided by the National Weather Service, was taken as a measure of surprise. Price change was measured by the average of the daily returns on the second- and the third-shortest maturity contracts (see Section I).

Table 5 presents the first results. The regressions there use the temperature forecast error as the dependent variable. The independent variables are the same day's OJ return plus the returns on two leading and two lagged days. (There is no causality implied or intended by choosing the "dependent" and "independent" variables in this way. Causality actually runs from weather to prices.) Results are given separately in Table 5 for the daily maximum and minimum temperatures, for each of the three available forecasts, and for observations weighted and unweighted by season.

TABLE 6—OJ FUTURES RETURNS AND TEMPERATURE FORECAST ERRORS WITH AGGREGATION OF LIMIT MOVES, OCTOBER 1975–DECEMBER 1981, OBSERVATIONS WEIGHTED BY SEASON

Hours Forecast is Ahead	Return Lag/Lead (Days)				
	b_{-2}	b_{-1}	b_0	b_{+1}	b_{+2}
Maximum (A.M.) Temperature Forecast					
36	.0692	.0671	−.102	.0449	−.0341
(1,257)	(1.46)	(1.25)	(−2.31)	(1.01)	(−.686)
24	.0654	−.00721	−.111	.0234	−.0545
(1,272)	(1.48)	(−.165)	(−2.74)	(.570)	(−1.33)
12	.0518	.0196	−.0121	.0482	−.0368
(1,263)	(1.30)	(.495)	(−.327)	(1.30)	(−.987)
Minimum (P.M.) Temperature Forecast					
36	.0542	−.101	−.236	.167	.0291
(1,272)	(.652)	(−1.23)	(−3.08)	(2.16)	(.377)
24	−.0955	−.00879	−.622	−.0395	−.00346
(1,263)	(−1.32)	(−.122)	(−9.25)	(−.584)	(−.0510)
12	−.00910	.0641	−.143	.0226	.106
(1,262)	(−.138)	(.981)	(−2.37)	(.375)	(1.74)

Notes: For regression and weights, see Table 5. All Durbin-Watson statistics were in the range 1.50–1.95. Adjusted R^2s were between 1 and 4 percent.

Given the preceding discussion, it might seem that the only relevant temperature observations would be for winter evenings (since freezes do not occur at other times); but the futures market deals in anticipations, so forecast errors during the morning hours or even errors during the summer months could conceivably contain meaningful information about the *probability* of a freeze later. The unweighted regressions with A.M. temperature errors do indeed contain some statistical significance. But the P.M. regressions weighted[17] by season are more significant. In the P.M. weighted cases, the contemporaneous OJ return is always statistically significant with the anticipated negative sign.

The P.M. temperature results indicate that the OJ futures price on a given day at the close of trading (2:45 P.M.) is a statistically significant predictor of the forecast *error* of the minimum temperature later that evening (from 7:00 P.M. until 7:00 A.M. the following morning). The price appears to be a slightly better predictor of the error in the forecast issued by the National Weather Service at 5:00 A.M. that same morning than of the errors made by the two other forecasts (5:00 P.M. the previous evening and 5:00 P.M. later the same day).

The futures price is not informationally efficient, however, because several later returns are statistically significant in some regressions. The significant negative coefficient b_{+1} in the P.M. 24-hour ahead case might be consistent with efficiency since trading ceases on day zero before the evening period begins and recommences on day +1 after the evening period ends (see Figure 1). However, the significant two-day later negative coefficients (b_{+2}) for the A.M. temperatures cannot be so easily dismissed.

There is ample a priori reason to suspect some effective informational inefficiency induced by limit move rules. There were 160 limit moves during the sample and prices on these days cannot reflect all information (see Section I). In a first attempt to eliminate this source of inefficiency, limit moves were "aggregated." The results are given in Table 6. For data used in this table, if a particular day registered a limit price move, the "eco-

[17]The weighting scheme is rather arbitrary but is was the only one I tried. January observations, in the middle of winter, receive the highest weight; July observations, in the middle of summer, receive the lowest. January observations are weighted seven times more heavily than July observations, intervening months are weighted linearly between January and July; i.e., February = 6, March = 5, ... June = 2, ... December = 6.

nomic" closing price for that day was assumed to be the price on the next subsequent day which did not have a limit move.

On Tuesday, January 6, 1976, for example, the March contract closed at 59.75 cents per pound. The next day registered a limit move of 3 cents; the reported closing price was 62.75 cents. On Thursday (January 8), which was not a limit move day, the settlement price was 64.4 cents. This was taken as an estimate of what the price would have been the preceding day (January 7) if the exchange had imposed no limits. Thus, the daily return for January 7 used in the regression was $\log_e(64.4/59.75) = 7.5$ percent. There was no observation used for January 8.

Limit moves often occur one after another. In such cases, the price on the first day with no limit move was brought back to the day of the first limit move and all intervening days were discarded.[18]

This procedure obviously overestimates the ability of the market to predict temperatures. Hindsight was used in that no one could know for sure on the first limit move day how many additional days with limit moves would follow. Thus, the results in Table 6 are biased in favor of finding market efficiency, as opposed to those in Table 5 that are biased against finding efficiency because of the exchange's own rules.

In Table 6, there is no longer a significant negative relation of temperature forecast error and later OJ returns. This indicates that the statistical significance of the lagging coefficients found in Table 5 was indeed due to the exchange's limit rules and not to some other possible source of informational inefficiency.[19] Notice that five of the six contemporaneous coefficients are significant and negative.[20]

To estimate the predictive content of OJ prices without resorting to hindsight, while at the same time including the extra information known to market participants that particular days had limit moves, the regressions in Table 7 were computed. A contemporaneous return and a lagged daily return were included as predictors along with slope dummies for limit move days.

Slope dummies are more appropriate than intercept dummies because the size of a limit move changed during the sample period (see Table 2).[21] Before January 1, 1979, the limit was 3 cents while it was 5 cents thereafter. As a consequence, only 39 out of 160 limit move days occurred during 1979–81 even though almost one-half of the sample observations were in those years. Thus, during 1979–81, the settlement price was more informationally efficient and the news that a particular day displayed a limit move constituted more material information. Slope dummies may not perfectly capture the greater importance of limit moves in the last three years of the sample, but at least they do weight these observations more heavily (by approximately 67 percent).

The F-statistics for these regressions indicate that the A.M. forecast errors cannot be

[18] If an up limit was followed by a down limit (or vice versa), day 1 was treated as if the return were zero and day 2 was discarded. The next included observation was then for day 3 (if it was not a limit move). In other words, for any sequence of limit moves followed immediately by another sequence in the opposite direction, the first closing price after reversal was brought back to the first day of the initial sequence. Then the price on the first day with no limit move is brought back to the first day of the second sequence.

[19] The one anomalous coefficient, b_{+1} in the 36-hour P.M. regression, has a positive sign. A single "significant" coefficient such as this is to be expected by chance among so many possibilities.

[20] The reader may notice that the number of observations differs by only one, 1263 to 1262, between the P.M. 24- and 12-hour regressions; yet the t-statistics on the contemporaneous returns are -9.25 and -2.37. Could this be caused by a single observation out of more than 1200? The answer is no. There are actually 138 observations that differed in these two regressions (due to missing data), but almost exactly one-half were missing from each regression. (There were other *common* missing observations.)

[21] Also, a slope dummy preserves the sign of the price change. This could be done, too, with intercept dummies, for example, using $+1$, 0, and -1 for up limit, normal, and down limit, but the slope dummy accomplishes this feat automatically while allowing for the nonstationarity in the size of a limit move.

TABLE 7—PREDICTIVE MODEL OF TEMPERATURE FORECAST ERRORS USING SLOPE DUMMY VARIABLES FOR LIMIT MOVE DAYS OCTOBER 1975–DECEMBER 1981, WEIGHTING BY SEASONS

Hours Forecast is Ahead	Contemporaneous		Lagged One Day		
	b_0	d_0	b_{-1}	d_{-1}	F^{a}
Maximum (A.M.) Temperature Forecast					
36	−.0636	−.0839	.0750	−.348	2.80
(1,391)	(−.495)	(−.475)	(.580)	(−1.91)	
24	.0992	−.213	−.0989	.0422	.897
(1,408)	(.835)	(−1.34)	(−.845)	(.254)	
12	.0198	−.0581	−.0807	−.0859	1.16
(1,400)	(.186)	(−.386)	(−.766)	(−.576)	
Minimum (P.M.) Temperature Forecast					
36	−.672	−.418	.0282	−.276	2.71
(1,407)	(−.329)	(−1.39)	(.131)	(−.898)	
24	.119	−1.55	.184	−.588	23.9
(1,399)	(.616)	(−5.82)	(.961)	(−2.17)	
12	−.119	−.643	.217	−.781	14.7
(1,398)	(−.697)	(−2.78)	(1.30)	(−3.32)	

Notes: The regression equation is $\log(A/F)_t = a + b_0R_t + d_0\delta_tR_t + b_{-1}R_{t-1} + d_{-1}\delta_{t-1}R_{t-1}$, where A is actual temperature, F is forecast temperature, R_t is return on day t, $\delta_t=1$ if there was a limit move on day t and zero otherwise.

See weighting scheme in Table 5.

T-statistics are shown in parentheses. Durbin-Watsons were in the range 1.59 to 1.99. Adjusted R^2s were in the range .0018 to .038.

[a]F-statistics for the regressors having no effect. The 95 percent fractile is approximately 5.6.

predicted by the current and lagged OJ returns plus a limit move slope dummy. This is also true for the P.M. 36-hour ahead forecasts. However, both the 24- and 12-hour ahead forecast errors can be improved by prior OJ returns.

The lack of predictive content of A.M. temperatures is, perhaps, not all that surprising because A.M. temperatures are relevant only to the extent that they predict freezes that evening. Apparently, this link is too weak to be picked up with statistical reliability by OJ returns.

The low predictive content for P.M. temperatures may be a disappointment until one reflects upon the scope of *possible* predictive ability. As shown in Table 3, about 90 percent of the variability in temperature is removed by the National Weather Service's forecast. The OJ prices predict a very small but still significant part of the remaining 10 percent.[22]

B. *Rainfall*

Orange juice prices are replotted in Figure 4 along with the day's total rainfall[23] (in tenths of inches) at Orlando. Unlike the earlier plot of price and temperature (Figure 3), no relation between the two series in Figure 4 is apparent to the naked eye.

The effect of rainfall on the crop is much less obvious than the effect of temperature. Most of the groves in Florida are not irrigated, so a long dry spell might be damag-

[22]It should be noted that all of the contemporaneous slope dummies (d_0) have negative signs. Also, the differences between the last two regressions in the table are intriguing but puzzling. The lagged slope dummy (d_{-1}) is more important for the 12-hour forecast error than for the 24-hour forecast error. Could this be related to the fact that the 12-hour forecast is not issued until after the market closes, while the 24-hour forecast is issued before it opens?

[23]Rainfall data are available only in the categories shown in Table 4. To construct Figure 4, the midpoint of each category was used as an estimate of the actual rainfall in inches. The A.M. and P.M. figures were added to obtain the total precipitation for the day.

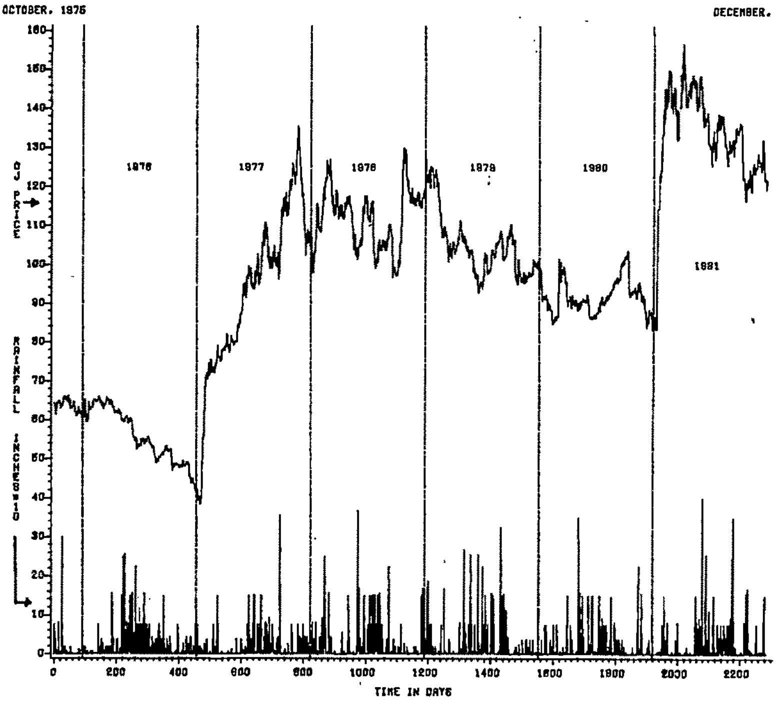


FIGURE 4. OJ FUTURES PRICES AND DAILY RAINFALL AT ORLANDO

ing. On the other hand, the crop could be reduced by extremely heavy rain or by wind damage from tropical storms (that appear in the rainfall time series because they also drop a lot of water).

For example, on November 6, 1981, the *Wall Street Journal* reported higher orange juice prices "...on news of a hurricane off the Florida coast," and on February 18, 1983, prices were purportedly higher due to "...talk of heavy rain." Some confusion about the effect of rainfall is disclosed in the latter story; it included a statement from the Florida Citrus Commission that the orange crop was "unscathed" by the rain. " 'Our oranges are enjoying the weather,' said a department spokesman, 'oranges need a lot of moisture.' " A commodities "analyst" stated that OJ traders drove up prices because they were confused by reports of rain damage to strawberries and tomatoes!

Whether or not the futures market understands the effect of rainfall is rather moot if the empiricist does not understand it well enough to develop a measure of rainfall surprise. With this admission in mind, let us plunge ahead into this turbid subject.

As shown previously in Section II (Table 4), National Weather Service rainfall forecasts are statistically significant but imperfect predictors of actual precipitation. I experimented with several different models of rainfall forecasts (including "probit" and logarithmic models), in order to find the most reliable predictor. It turned out that the largest reduction in variance was obtained with the simplest of regression models,

$$A_t = a + bF_t,$$

where $A_t = 1,\dots,9$ is the actual rainfall by category on day t and F_t is the forecast "probability of rain." The adjusted R^2 of this regression ranged between .118 and .332 (see Table 8). It is interesting to note that predictive ability for rainfall rises more rapidly as the prediction period approaches than it does in the case of temperature (compare Table 3).

Table 8 contains F-statistics from regressions relating the rainfall forecast error to

TABLE 8—PREDICTIVE MODEL OF RAINFALL FORECAST ERRORS USING SLOPE DUMMY VARIABLES FOR LIMIT MOVE DAYS OCTOBER 1975–DECEMBER 1981, NO WEIGHTING

Hours Forecast is Ahead	Adjusted R^2 of Weather Service Forecast[a]	F-Statistic of OJ Return Predictive Power[b,c]
A.M. Rainfall		
36 (1,371)	.239	.362
24 (1,393)	.265	.410
12 (1,372)	.332	.417
P.M. Rainfall		
36 (1,393)	.118	.388
24 (1,374)	.165	.230
12 (1,384)	.225	.629

[a]Actual rainfall A_t by category, ($A_t = 1,2,\dots 9$), was predicted by the Weather Service's "probability of rain," F_t, in the simple regression model $A_t = \hat{a} + \hat{b}F_t + \varepsilon_t$; the forecast error ε_t was then used as the dependent variable in another regression model with OJ returns as predictors (see fn. c below).

[b]The 95 percent fractile of the F-statistic is approximately 5.6.

[c]The regression model was $\varepsilon_t = a + b_0 R_t + d_0\delta_t R_t + b_{-1}R_{t-1} + d_{-1}\delta_{t-1}R_{t-1}$, where ε_t is the Weather Service's rainfall prediction error, R_t is the OJ return on day t and δ_t is $+1$ if day t had a limit move, otherwise zero. No coefficient was significant and coefficients are not reported for reasons of space.

the contemporaneous and lagged OJ return plus a slope dummy for limit moves, that is, the same purely predictive model as the one for temperature in Table 7. As might have been anticipated in light of the preceding discussion, OJ returns appear to have no significant predictive power for rainfall.[24] There was not a single significant coefficient out of the 24 possible and no F-statistic is significant in any of the six regressions.

[24]A similar model was computed with a dependent variable defined as the *absolute value* of the rainfall forecast's prediction error. Of course, this would not be a legitimate model from an efficient markets perspective since it would not imply predictive ability of the direction of error (even if it had worked). It is, however, suggested by the possibility that either too much or too little rain is bad for the orange crop. As it turned out, the model had even lower explained variance than the model in Table 8 which preserved the sign of the rainfall prediction error.

C. *Nonweather Influences on OJ Prices*

The small predictive power for temperature and rainfall seems to imply that influences other than weather are affecting OJ returns. What might they be? In an attempt to find out, news stories in the financial press were systematically examined.

From October 1, 1975 through December 31, 1981 (the sample period of the paper), a total of 91 articles related to oranges appeared in the *Wall Street Journal*; 26 articles reported either results of weather (17) or forecasts of weather (9). Of the 26 weather articles, 25 concerned temperature and 1 concerned rainfall. There were 22 articles disclosing crop forecasts by the U.S. Department of Agriculture, 15 articles reporting price movements with no explanation, 7 articles about international conditions (Canadian and Japanese imports and Brazilian exports), 6 articles about supermarket supplies, and 15 miscellaneous articles. In this last category, the subjects ranged from product quality (4) and new products (1) through antitrust action against the Sunkist cooperative in California (3), to such truly unclassifiable stories as orange rustlers in Florida and advertising contracts with Anita Bryant.

The number and content of weather stories shows that weather is considered important and that rainfall is a relatively minor factor compared to temperature. Among the other topics, *ex post* stories about futures price movements per se and most of the miscellaneous stories could not possibly have been about true influences on earlier OJ price variation. Agricultural crop forecasts, though, would seem likely to have moved prices in some direction. Perhaps international news, reports of supermarket supplies, and antitrust actions are also relevant. The variability of returns was computed for periods ending on the *Wall Street Journal* publication date of such articles and including two prior trading days (to allow for news leakage). This variability is compared in Table 9 to the variability of returns on dates with no orange juice news.

TABLE 9—VARIABILITY OF OJ FUTURES RETURNS ON DAYS WITH NEWS ABOUT ORANGE JUICE IN THE *WALL STREET JOURNAL*, OCTOBER 1975–DECEMBER 1981

	No News (1)	Weather (2)	Crop Forecast (3)	Supplies, Antitrust, International (4)	Miscellaneous (5)
Standard Deviation of Returns	1.53 (1361)	2.86 (64)	2.01 (60)	1.97 (34)	1.37 (34)

	Comparisons Among	*F*-Statistic	Tail Probability
Levene's Test for Equal Variances[a]	Cols. (1)–(5)	22.5	0.0000
	Cols. (1), (3), (4), (5)	9.83	.0018
	Cols. (2), (3), (4).	8.99	.0033

Notes: Standard deviation of returns are shown in percent per day, with sample size shown in parentheses; returns on an equally weighted index of the second and third from the shortest maturity contracts on the day of the news story and on the two preceding trading days.

Sample sizes are smaller than the number of possible days because of overlapping dates among articles. For overlapping dates, returns were assigned hierarchically to category (2) (Weather) first, then to categories (3), (4), and (5), respectively.

[a]See Brown and Forsythe.

The miscellaneous category has a low volatility. It is even lower than the variability of returns on days with no news stories. Volatility of returns is highest during periods when stories about weather were published. During periods associated with stories about crop forecasts, retail supplies, antitrust actions, and international events, volatility is higher than during "no news" periods. However, it is significantly *lower* than during periods with weather-related news stories.

From this evidence, weather remains as the most important identifiable factor influencing OJ returns. Crop forecasts and other newsworthy events have an influence, but their frequency is too small and their impact too slight to explain a material part of the variability in returns left unexplained by weather. As Table 9 shows, there is substantial volatility (a daily standard deviation of returns of 1.53 percent per day), on days that are not associated with *any* story about oranges in the *Wall Street Journal*; and these days constitute about 87 percent of the sample observations.

In addition to events important enough to appear in special orange juice stories in the financial press, other influences on supply and demand might be directly measurable. For instance, stock market returns could measure general economic activity and thus provide a proxy for consumer demand. Canada is the largest customer for U.S. orange juice, so the Canadian dollar/U.S. dollar exchange rate might have a measurable impact on orange juice because it would proxy for Canadian demand. Energy prices could affect short-term supply because they influence the cost of operating farm equipment and the costs of processing and distributing the product. Petroleum is also a direct ingredient of fertilizer and a major component of fertilizer production costs.

Table 10 offers evidence about the influence of these and other variables on OJ price movements. Two regressions were computed. The first involves the OJ return as dependent variable. It shows that cold temperatures indeed cause OJ price movements, but general stock market returns, changes in the Canadian dollar exchange rate, and oil stock returns (a measure of energy prices), have no significant influence.

The second regression in Table 10 uses the squared OJ return as dependent variable. This was done because the objective here is merely to identify sources of price movements in either direction, as opposed to test-

TABLE 10—T-STATISTICS OF EXPLANATORY FACTORS FOR OJ RETURNS, NO CONSIDERATION OF LIMIT MOVES, DAILY DATA, OCTOBER 1975–DECEMBER 1981

	Dependent Variable	
Explanatory Variable	OJ Return	Squared OJ Return
Max $(32-T_{-1},0)$[a]	5.40	7.99
Max $(32-T_{-0},0)$	3.69	8.09
(Oil Stock Return)$_{-1}$[b]	−.618	.385[g]
(Oil Stock Return)$_0$	.624	2.11[g]
(VW Market Return)$_{-1}$[c]	.525	−1.05[g]
(VW Market Return)$_0$	−.120	−1.53[g]
(Δ CDN exch. Rate)$_{-1}$[d]	−.417	−.759[g]
(Δ CDN exch. Rate)$_0$	.577	.938[g]
Monday[e]	−2.18	4.23
Weather-Related News Story[f]	–	9.36
Crop Forecast News Story[f]	–	3.35
Supplies or Int'l News Story[f]	–	−.563
Miscellaneous News Story[f]	–	−1.47
Multiple Adjusted R^2	.0668	.268
F-Statistic for Regression	13.4	45.0
Durbin-Watson	1.81	1.39
Number of Observations	1,559	1,559

[a] T_t is the minimum temperature at Orlando on day t.
[b] Return on an equally weighted portfolio of oil stocks listed on the NYSE and the AMEX, consisting of up to 45 firms. The sample consisted of all listed oil firms covered in the 1982 *Value Line* service.
[c] Value-weighted index of all NYSE and AMEX stocks.
[d] Percentage change in the Canadian/U.S. dollar exchange rate.
[e] Dummy variable; 1 if Monday, 0 otherwise.
[f] Dummy variable; 1 if news story in this category in the *Wall Street Journal* on day t or $t+1$, zero otherwise.
[g] T-statistic for the squared explanatory variable.

ing the direction of influence of particular variables. Using the squared return permits the inclusion of dummy variables on news story dates without having to decide whether the story should be associated with a positive or negative price change. To illustrate the problem, take the case of crop forecast stories. It would be very hard to know whether a particular forecast by the Department of Agriculture is above or below the previously expected production level without looking at the OJ price movement itself.

In this second regression, cold weather remains very significant and stories related to weather and to crop forecasts are significant as well (the latter result confirms the implications drawn from Table 9). The contemporaneous squared oil stock return is also significant, though its t-statistic indicates a much lower level of influence. (This is something of a curiosity in that oil stock returns are unrelated in direction to OJ returns in the first regression.) Finally, notice that only 27 percent of the variability in squared OJ returns is explained by all of these variables combined. Most of the variability remains unexplained.[25]

D. *Supply Shocks vs. Demand Shocks*

Variability in OJ prices could be caused by shifts in demand induced by changes in the prices of substitute products. The prices of apple juice, tomato juice, and soft drinks, inter alia, should influence the demand for orange juice. We have seen already in Table 10 that general consumer demand and the demand of the largest foreign customer (Canada) are not important relative to the supply shocks of weather, energy prices, and crop forecasts. Table 11 provides information about the relative importance of more micro demand shocks.

For firms in the orange juice business and for certain firms producing substitutes, daily stock returns were related, firm by firm, to OJ returns. In each case, the firm's return was regressed on the contemporaneous OJ return, plus two leading and two lagged OJ returns, plus slope dummies for limit move days on the OJ exchange. The F-statistics of the regression were examined for significance. In cases where significance was indicated, the coefficients were examined for direction of comovement between equity and OJ returns.

Two basic types of firms were examined. The first type consists of firms whose SIC (standard industrial classification) code on the CRSP tape indicated that it was in some aspect of the orange juice or a related food-processing business. (It had the same SIC

[25] These regressions are obviously misspecified (for example, notice the Durbin-Watson statistics in the second regression). However, they are intended merely to characterize the data, not to test any particular theory, so it seems doubtful that much can be learned by using more sophisticated econometric methods.

TABLE 11—RETURNS ON AGRICULTURE RELATED EQUITIES AND RETURNS ON ORANGE JUICE FUTURES[a]

Company[b]	Line of Business	Relation to OJ Returns[c]
American Agronomics	Owns 9200 acres of Fl. citrus; Produces and markets OJ	None (+)
CHB Foods	Produces and markets pet food, fish, vegetables and fruit	None
Castle & Cooke	Produces and markets pineapples, bananas, fish, broccoli, sugar; Owns Hawaii land	Positive
Consolidated Foods	Manufactures and distributes coffee, candy, sugar, soft drinks	Positive
Curtice-Burns	Processes and packs fruits and vegetables, soft drinks, Mexican food, frozen vegetables	None
Del Monte[d]	Produces fresh bananas and pineapples; processes seafood	None
Di Giorgio	Diversified food processor including citrus, Italian food, sells OJ in Europe; Has some Fl. land	None
Green Giant[d]	Produces canned and frozen vegetables	None
Norton Simon	Produces tomato-based food products, popcorn, cooking oil, liquor	None (−)
Orange-Co. Inc.	Owns 8100 acres of Fl. citrus; Produces and markets OJ	None
J. M. Smucker	Produces jellies, condiments, syrups, and canned fruit drinks	None (−)
Stokeley Van Camp	Produces Gatorade and canned and frozen vegetables	None
Tropicana[d]	Processes citrus juice; Owns a few Fl. groves which are experimental plantings	Negative
United Foods	Produces frozen vegetables	None

[a]Equities with the standard industrial classification of food manufacturers and processors with the same four-digit SIC codes as Di Giorgio, Orange-Co. or Tropicana, and with at least 100 daily return observations in the period October 1975–December 1981.

[b]In addition to these companies, regressions were also run with soft drink equities, Coca-Cola, Dr. Pepper, MEI, Pepsi Cola, and Royal Crown. None of these regressions were significant.

[c]"Positive" or "Negative" indicates that the regression's F-statistic was significant at the 5 percent level. The regression's dependent variable is the equity's return and independent variables are two leading, contemporaneous, and two lagged orange juice futures returns plus corresponding slope dummies for limit moves. A symbol in parentheses indicates a marginally significant regression (at the 10 percent level).

[d]Companies no longer listed on the New York or American Exchange.

code as Di Giorgio, Tropicana, or Orange-Co., three companies known in advance to be in the orange juice business.) All such companies are listed by name in Table 11.

The second type of company produced soft drinks (see Table 11, fn. b). No soft drink producer had a significant relation to orange juice. So changes in OJ demand due to changes in soft drink prices are not revealed in the data.[26]

Turning back to the first type of firm, Table 11 indicates that many were not related to OJ prices. This was true even for such companies as Orange-Co., whose principal business is growing oranges and producing juice. There are several possible explanations for the lack of significant comovement in such a firm. First, consider the impact of supply shocks: an increase in OJ prices due to, say, cold weather, would not affect the firm if the gain in the value of its Florida land were offset by a reduction in the value of its processing and distribution divisions, or if the firm had hedged its own supply by selling OJ futures.

A demand shock, however, should affect the firm unequivocally unless it overhedged in the futures market. For example, an exogenous increase in OJ demand raises the value of its land and, if there are fixed costs, also raises the value of its production and distribution facilities. Thus, the lack of significant comovement between OJ prices and firms such as Orange-Co., Di Giorgio, and Amer-

[26]One of these companies, Coca-Cola, also produces orange juice, so a lack of comovement due to shifts in prices of orange juice substitutes might be expected for this particular firm; roughly, what it gains in the soft drink business might be lost in the orange juice business, or vice versa.

ican Agronomics, who grow *and* process juice, suggests that most of the OJ price volatility is due to supply shocks instead of demand shocks.

This is reinforced by the case of Tropicana, a processor owning virtually no land. It is the only such firm and also the only firm whose equity comoves negatively and significantly with OJ prices. It is conceivable, of course, that this negative relation is induced by a combination of demand shocks and Tropicana purchasing too many futures contracts (more than its own anticipated requirements), but it seems more plausible that the relation is induced directly by supply shocks that squeeze Tropicana's profit margin.

Two companies, Castle & Cooke and Consolidated Foods, produce OJ substitutes and have positive comovement with OJ prices (as is expected if OJ prices move because of supply shocks). One firm, Smucker, buys oranges for jam and has a marginally negative comovement (also explainable by OJ supply shocks). The only anomalous firm is Norton Simon, a producer of substitutes such as tomato juice and liquor (but its negative comovement is of only marginal significance). Some wits have suggested that Norton-Simon actually produces a complement, not a substitute, product. Vodka, one of its biggest sellers, is often consumed with orange juice.

Overall, the evidence in Table 11 supports the view that supply shocks are the principal cause of OJ price movements. Unfortunately, the identity of such shocks remains at least a partial mystery. Weather is important, but measured weather explains only a small fraction of the volatility in OJ prices.

III. Summary and Conclusion

The market price of frozen concentrated orange juice is affected by the weather, particularly by cold temperatures. A statistically significant relation was found between OJ returns and subsequent errors in temperature forecasts issued by the National Weather Service for the central Florida region where most juice oranges are grown. Orange juice prices are much less related to errors in rainfall prediction. Indeed, no significant statistical association was found between these variables.

The OJ futures price is rendered informationally inefficient by the existence of exchange-imposed limits on price movements. This inefficiency manifests itself in the data by allowing temperature surprises to have apparent predictive power for *later* price changes. When limit moves are taken into account, however, temperature has no remaining predictive content.

There is, nevertheless, a puzzle in the OJ futures market. Even though weather is the most obvious and significant influence on the orange crop, weather surprises explain only a small fraction of the observed variability in futures prices. The importance of weather is confirmed by the fact that it is the most frequent topic of stories concerning oranges in the financial press and by the ancillary fact that other topics are associated with even less price variability than is weather.

Possible sources of orange juice demand and supply movements such as substitute product prices, general demand, export demand, and production costs were also examined here. Yet *no* factor was identified that can explain more than a small part of the daily price movement in orange juice futures. There is a large amount of inexplicable price volatility.

REFERENCES

Blume, Marshall, E., and Stambaugh, Robert F., "Biases in Computed Returns: An Application to the Size Effect," *Journal of Financial Economics*, November 1983, *12*, 387–404.

Brown, Morton B. and Forsythe, Alan B., "Robust Tests for the Equality of Variances," *Journal of the American Statistical Association*, June 1974, *69*, 364–67.

Dimson, Elroy, "Risk Measurement When Shares are Subject to Infrequent Trading," *Journal of Financial Economics*, June 1979, *7*, 197–226.

French, Kenneth R., "Stock Returns and the Weekend Effect," *Journal of Financial Economics*, March 1980, *8*, 55–69.

Gibbons, Michael R. and Hess, Patrick, "Day of the Week Effects and Asset Returns,"

Journal of Business, October 1981, *54*, 579–96.

Hopkins, James T., *Fifty Years of Citrus: The Florida Citrus Exchange: 1909–1959*, Gainesville: University of Florida Press, 1960.

Keim, Donald B., "Size Related Anomalies and Stock Return Seasonality: Further Empirical Evidence," *Journal of Financial Economics*, June 1983, *12*, 13–32.

Mandelbrot, Benoit, "Forecasts of Future Prices, Unbiased Markets, and 'Martingale' Models," *Journal of Business*, January 1966, *39*, 242–55.

McPhee, John, *Oranges*, New York: Farrar, Straus, and Giroux, 1967.

Scholes, Myron and Williams, Joseph, "Estimating Betas from Nonsynchronous Data," *Journal of Financial Economics*, December 1977, *5*, 309–27.

Theil, Henri, *Applied Economic Forecasting*, Amsterdam: North-Holland, 1966.

Citrus Associates of the New York Cotton Exchange, *Citrus Futures*, Four World Trade Center, NY 10048, undated.

Florida Department of Agriculture, *Florida Agricultural Statistics Summary*, Tallahassee: Florida Department of Agriculture, various years.

U.S. Department of Agriculture, *Agricultural Statistics*, Washington: USGPO, various years.

Prelude to Macroeconomics

By RICHARD STARTZ*

Two schools of macroeconomic thought compete today. The Keynesian school attempts to analyze each sector of the economy using the usual tools of optimizing models, but produces general equilibrium descriptions of a macroeconomy which are rarely Pareto efficient. For this reason, economists question the internal consistency of Keynesian models. In contrast, the neoclassical or "rational expectations" school maintains consistency with the principles of perfect competition and flexible prices. In essence, the neoclassicists' macroeconomy behaves as an Arrow-Debreu general equilibrium. Once this is understood, we realize that the economy is Pareto efficient, though this does not preclude occasional *ex post* bad draws. Government policy can be expected to be either neutral or damaging. (In fairness, I am describing polar cases of the Keynesian and neoclassical view.)

Nearly all economists are extraordinarily prejudiced in favor of models exhibiting rational behavior (as am I) and the last decade has seen an almost complete intellectual victory for the neoclassical school. Complete victory has been elusive for a single reason. In apparent ignorance of the intellectual arguments of the neoclassical school, the economy persists in behaving pretty much as the modern Keynesian models predict. As premier examples, neither the Great Depression nor the recent massive recession was (in my opinion) a Pareto-efficient equilibrium.

The model I present below rigorously adheres to the rule that agents should follow rational principles. In this paper that rule means that agents equate marginal rates of substitution to relative prices. At the same time, I insert a single piece of imperfect information which prevents the formation of a complete Arrow-Debreu general equilibrium. The model predicts qualitative behavior of *GNP* and employment which is analogous to Keynesian predictions. Government spending is shown to increase *GNP* and economic welfare.

The substantive results of the paper appear in the next three sections. Section I presents the role of imperfect information in the labor market and then goes on to solve for general equilibrium in the absence of government intervention. Section II examines the Keynesian-like behavior of this equilibrium. Section III examines the impact on *GNP*, aggregate labor supply, and welfare of balanced-budget government spending. The model produces four major results.

1) Aggregate spending and labor supply decisions are not simply the sum of individual decisions. The model identifies the logical error that Paul Samuelson has labeled the "fallacy of composition."

2) Say's Law fails. A unit increase in aggregate supply produces a less than unit increase in demand for output.

3) An increase in government spending increases *GNP* and reduces unemployment.

4) An increase in government spending can generate a Pareto improvement in individual utility.

The last section of the paper discusses some of the ways this model differs from the way we usually think about the economy. While the paper develops a particular model of aggregate demand, its real goal is to demonstrate a general principle: once the right set of mathematics is put together, it is easy to produce a model of the economy which is at once rational and Keynesian. In the specific setting I present, all the results fol-

*Department of Economics, University of Washington, Seattle, WA 98195. The first version of this paper was written while visiting at the Graduate School of Business, Stanford University. The final version was written while a member of the Finance Department, University of Pennsylvania. Shelly Lundberg deserves many thanks for extensive aid and partial absolution from any remaining errors. Stanley Fischer, Mark Flannery, Robert Solow, a number of other friends and colleagues, and two anonymous referees all contributed valuable constructive criticism for which I am most grateful.

low from a single "problem," which is located in the labor market, not the market for goods and services. The model is very spare, abstracting from many of the details important to a full understanding of the macroeconomy. This spareness clarifies the global claim of the paper—that rational and Keynesian explanations are complementary, not contradictory.

I. The Model

I begin with the labor supply/goods consumption problem faced by the individual worker. All the results of the model follow from a single assumption—that a worker's productivity is imperfectly observed. I assume that an employer observes hours worked and many other signals of the productive output of a worker. Also, an employer can observe total output of a group of workers exactly. Nonetheless, an individual worker's productivity is measured only with error. One could think of the worker as supplying labor hours adjusted for imperfectly observable "effort," or that the effectiveness of the hours depend on human capital which the worker has acquired but cannot perfectly signal. In any event, I require only that labor productivity be imperfectly observed and that there is increasing disutility to providing this effective labor. I label the true value of the worker's labor supply (measured in marginal product units) l and label the signaled value $l+\varepsilon$. Think of ε as observational error; just random noise in the system. By assumption, the worker knows ε. Employers naturally try to infer ε from whatever information is available. Understand that there is no mechanism to measure an individual's ε, neither *ex ante* nor *ex post*. Therefore, labor contracts cannot include clauses or actions requiring its revelation.

The employer infers each worker's marginal product from the signal and pays each worker accordingly. Each employer has enough workers so that the law of large numbers applies, so per capita production is effectively riskless. (Alternatively, individual noise might be thought of as perfectly diversifiable risk.) Workers face the wage offer schedule with certainty. Thus there is no interesting (read "priced") risk in this model. Competition forces employers to pay a worker the expected value of his or her marginal product, where the expectation is conditioned on the signal $l+\varepsilon$, so firms make zero economic profits. Workers receive $w\cdot E(l|l+\varepsilon)$.

Assume that l and ε have a joint normal distribution and that the unconditional mean of ε is zero. As is usual in rational expectations models, it is assumed that the moments of the relevant statistical distribution are all common knowledge. The conditional expectation of joint normal variables is an easily calculated linear function. (These assumptions are consistent with the model that follows. The linearity they imply is of considerable mathematical convenience.) The conditional expectation of l is given by

$$(1)\quad E(l|l+\varepsilon)=(1-\beta)L+\beta(l+\varepsilon),$$

where L is the population mean of l and β is the population regression coefficient of l on $l+\varepsilon$. The exact value of β, where $\beta=\mathrm{cov}(l,l+\varepsilon)/\mathrm{var}(l+\varepsilon)$, depends on the cross-sectional distribution of endowments and tastes; β is less than 1 and greater than 0. (The latter requires that l and ε are not too negatively correlated, which I assume to be the case.)[1]

Figure 1 shows the wage function faced by a typical individual. On average, workers are paid just what they are worth. However, an individual who supplies greater than average l is undercompensated (and vice versa for below average l) because employers can only partially distinguish high l from high ε. If an individual increases his or her effective labor supply one unit, social production rises one unit. The individual's compensation rises by only the fraction β of the unit. This identifies the deviation between private and social marginal conditions which is responsible for

[1] The value l is an endogenous variable, so one must compute out the equilibrium in order to calculate β. However, to provide an illustration, assume l and ε are uncorrelated. Then, $\beta=\mathrm{var}(l)/[\mathrm{var}(l)+\mathrm{var}(\varepsilon)]$. With no noise the signal is perfect, $\beta=1$; with infinite noise the signal is worthless, $\beta=0$.

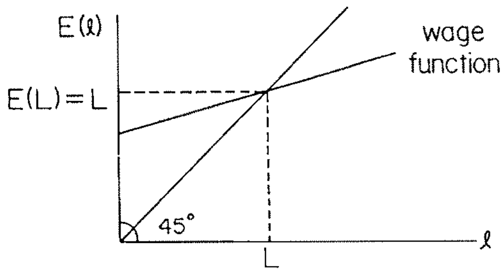


FIGURE 1

the results of the model. This deviation is solely responsible for the failure of an Arrow-Debreu general equilibrium. It may help to view the role played by aggregate labor supply by analogy to a classic "public good." Each individual properly regards L as exogenous with respect to private decisions, even though as a group individuals may agree that it would be better if L were higher.

The alert reader will recognize this particular piece of mathematics from the 1983 paper by Shelly Lundberg and myself. Both that paper and this focus on deviation of private and social marginal conditions. This similarity is not accidental. It was our intent at the time that our discrimination paper could serve as a precursor for research into macroeconomic phenomena.

A. *The General Equilibrium Model*

I derive the general equilibrium model in three logically separate steps. First, I find the demand functions for an individual who maximizes utility, taking market equilibrium as fixed. Second, these individual demand schedules are aggregated into market demand functions. Third, general equilibrium based on the market demand functions is calculated. To emphasize the difference between individual and aggregate demands, I use lowercase symbols for individuals and uppercase symbols for per capita aggregate values. Thus, l is the labor supplied by one individual and $L=\Sigma l/n$ is the per capita aggregate labor supply.

Each individual consumes two consumption goods, c and g, and supplies labor, l. Later, I will associate c with private consumption and g with government spending, but as there are no public good aspects to g, we actually just have two goods, valued for their private utility, that are imperfect substitutes. Individuals have identical utility functions, but differ in endowments and noise (ε) levels. An individual's endowment of effective labor is l^*. Individuals also own fixed capital endowments, k, which they rent to firms.

Production is characterized by competitive firms with linear production functions; c and g are assumed to be perfect substitutes in production. The common nominal price of c and g is unity; the real wage rate is w, and the real rental rate on capital is q. The per capita production function is

$$(2) \qquad C+G=a_L L+a_K K.$$

My model, as with most macro models, has little to say about relative prices. The linear production function serves to fix prices according to conditions of constant marginal productivity. The real wage rate is set at $w=a_L$ and the real return to capital at $q=a_K$. (Note that GNP is easily measured as $C+G$.) It should be evident from the outset that since relative prices are fixed, none of the results of the model derive from changes in relative prices.

Individuals have Stone-Geary utility over c, g, and leisure. "Budget share" parameters α_1, α_2, and $(1-\alpha_1-\alpha_2)$ are assumed to be between zero and one. (α_1 is a budget share in the sense that it gives the ratio of expenditure on C above the subsistence level divided by full income.) Subsistence requirements for the three commodities are γ_1, γ_2, and γ_3. An individual solves the following maximization problem (under certainty):

$$(3) \qquad \max_{c,g,l} U=\alpha_1\ln(c-\gamma_1)+\alpha_2\ln(g-\gamma_2) + (1-\alpha_1-\alpha_2)\ln(l^*-l-\gamma_3)$$

$$\text{subject to} \quad c+g=wE(l|l+\varepsilon)+qk.$$

Solving for individual demands is straightforward. The demand functions show three

slight differences from the usual ones. First, the effective marginal wage is $w\beta$, rather than w. Second, L becomes part of the individual's endowment. Third, the exogenous ability to fool employers into thinking one is a hard worker, $\varepsilon > 0$, adds to an individual's endowment. Let f_I be an individual's full income. (Full income is labor income plus capital income plus the market value of leisure, less the cost at market prices of the subsistance levels of c, g, and leisure.) The individual demands for c, g, and leisure are given by

$$(4) \qquad f_I = w[(1-\beta)L + \beta(l+\varepsilon)] + qk + w\beta(l^* - l) - \gamma_1 - \gamma_2 - w\beta\gamma_3$$

$$c - \gamma_1 = \alpha_1 f_I$$

$$g - \gamma_2 = \alpha_2 f_I$$

$$l^* - l - \gamma_3 = (1 - \alpha_1 - \alpha_2)\frac{1}{w\beta} f_I.$$

These demand schedules just say that α_1, α_2, and $(1-\alpha_1-\alpha_2)$ give real budget shares of individual full income evaluated at market prices. (The derivation of the demand schedules contains one minor inconsistency. Having assumed that the ε are normally distributed, it is necessarily true that a finite fraction of the population has income below the subsistence level. The observed ε would be truncated normal. The truncation bias can be made arbitrarily small by assuming that mean income is large relative to the standard deviation of ε. I ignore this truncation problem throughout.)

Microeconomic demand schedules emphasize prices, endowments, and tastes, as in equations (4). In the macroeconomic tradition, greater emphasis is given to the role of aggregate income in determining consumption possibilities and less weight is placed on the disaggregation of demands among different products. Equation (5) shows an individual's total demand for goods. Individual demand depends on aggregate variables, Y and K, as well as individual characteristics. (In transforming (4) into (5), I've skipped ahead and employed one of the obvious characteristics of the general equilibrium, $Y = wL + qK$.) Adding demands for c and g from equation (4) and substituting in for f_I, we show

$$(5) \quad \text{individual demand for goods} = \gamma_1 + \gamma_2 + (\alpha_1+\alpha_2)\{-(\gamma_1+\gamma_2+w\beta\gamma_3) + (1-\beta)[Y - qK] + qk + w\beta[l^* + \varepsilon]\}.$$

Note that the marginal propensity to spend out of aggregate income (the coefficient on Y, treating individual endowments as constant), $(\alpha_1+\alpha_2)(1-\beta)$, depends on tastes for goods versus leisure, as well as β. It is also useful to look at the demand for goods plus the value of leisure, equation (6), with leisure valued at the marginal wage rate $w\beta$.

$$(6) \quad \text{individual demand for goods and leisure} = (1-\beta)[Y - qK] + qk + w\beta[l^* + \varepsilon].$$

The per capita aggregate demand schedule is found by adding up demand schedules for all individuals and dividing by the size of the population. Equivalently, aggregate demand (AD), equation (7), and Net Economic Welfare (NEW),[2] equation (8), are computed as traditional macroeconomic functions of aggregate income by taking the mean values of equations (5) and (6), respectively,

$$(7) \quad AD = \gamma_1 + \gamma_2 + (\alpha_1+\alpha_2)\{-(\gamma_1+\gamma_2+w\beta\gamma_3) + (1-\beta)Y + \beta[qK + wL^*]\};$$

$$(8) \quad NEW = (1-\beta)Y + \beta[qK + wL^*].$$

Traditionalists will note the close analogy between equation (7) and the Samuelson 45° line model in that aggregate demand depends explicitly on aggregate income, as well

[2] NEW is a better measure of welfare than is GNP, as NEW accounts for substitution between goods and leisure. Specifically, a one dollar increase in NEW is equivalent to a welfare improvement of one dollar times the marginal utility of income.

as endowments. (I briefly explore this analogy in Section IV.)

There are two ways to go about solving for market equilibrium. In the first, the microeconomic approach to general equilibrium, we aggregate equations (4), noting that L equals mean l and mean ε equals zero. Equations (9) show market demands for goods and leisure in terms of mean full income, F_I:

$$F_I = w(L^* - \gamma_3) + qK - \gamma_1 - \gamma_2 \tag{9}$$

$$C - \gamma_1 = \frac{\alpha_1 \beta}{1-(1-\beta)(\alpha_1 + \alpha_2)} F_I$$

$$G - \gamma_2 = \frac{\alpha_2 \beta}{1-(1-\beta)(\alpha_1 + \alpha_2)} F_I$$

$$L^* - L - \gamma_3 = \frac{(1-\alpha_1 - \alpha_2)}{1-(1-\beta)(\alpha_1 + \alpha_2)} \frac{1}{w} F_I.$$

In the second approach, the macroeconomic approach to general equilibrium, we set aggregate demand equal to aggregate income. Equation (10), the reduced form of equation (7), shows GNP; equation (11), the reduced form of equation (8), shows NEW:

$$GNP = \gamma_1 + \gamma_2 + \frac{(\alpha_1 + \alpha_2)\beta}{1-(1-\beta)(\alpha_1 + \alpha_2)} F_I; \tag{10}$$

$$NEW = \gamma_1 + \gamma_2 + w\beta\gamma_3 + \frac{\beta}{1-(1-\beta)(\alpha_1 + \alpha_2)} F_I. \tag{11}$$

Naturally, the two approaches arrive at the same result. Note that consumption of each good, the supply of labor, and the utility of the mean individual are all increasing functions of β. With $\beta < 1$, GNP and NEW are lower than in a first-best, $\beta = 1$ world. While sad, this unfortunate state is not the basis for my claim that Say's Law fails to hold. It is unfortunate that information about labor productivity is limited. It is also unfortunate that we are not all wealthier in other ways. I do not regard these as macroeconomic "failures."

II. The Fallacy of Composition and the Failure of Say's Law

In this section I examine Samuelson's "fallacy of composition" and the failure of Say's Law by considering first a decrease in desired leisure and then a rain of capital equipment. The key in both cases lies in comparison of individual and aggregate demand.

A. *The Fallacy of Composition*

Suppose everyone changes preferences in favor of more work and less leisure. Specifically, consider a uniform unit decrease in γ_3. First, we look at the reaction of each individual in isolation from the market, as in equations (5) and (6). Secondly, we look at changes in market equilibrium, as in (10) and (11).

Individual response to a declining taste for leisure is just what one would expect in an equilibrium where agents have equated marginal rates of substitution with relative prices. Individual full income, f_I, rises by one unit times the marginal wage rate, $w\beta$. Demand for goods, in (5), rises by $(\alpha_1 + \alpha_2)$, the budget share of goods in full income, times $w\beta$. Total demand for goods and leisure, in (6), is unchanged, as agents have merely substituted less leisure for more goods.

Aggregate response differs from a mere summation of individual responses because aggregate income and labor supply change, because aggregate variables are properly regarded as exogenous by individuals but are endogenous in market equilibrium. This difference can be seen as a "multiplier" effect in equations (7) and (8). Alternatively, aggregate full income, F_I, rises by one unit times the marginal product of labor, w, and GNP, in (10), rises accordingly. Equilibrium GNP rises by more than one would expect from looking at individual goods demand, where the individual demands were evaluated taking aggregate income as fixed. Equation (12) compares the change in (per capita)

aggregate *GNP* to the change in individual goods demand.

$$\text{(12)} \quad \left.\frac{dGNP}{d(-\gamma_3)}\right|_{\text{aggregate}}$$

$$= \frac{1}{1-(1-\beta)(\alpha_1+\alpha_2)} \cdot (\alpha_1+\alpha_2)w\beta$$

$$> (\alpha_1+\alpha_2)w\beta = \left.\frac{dGNP}{d(-\gamma_3)}\right|_{\text{individual}}.$$

Contrary to the measured effect on welfare predicted looking at individual behavior in isolation, *NEW* rises in equilibrium.[3]

$$\text{(13)} \quad \frac{dNEW}{d(-\gamma_3)} = \frac{(1-\beta)(\alpha_1+\alpha_2)}{1-(1-\beta)(\alpha_1+\alpha_2)} w\beta > 0.$$

Notice two points which justify labeling as "macroeconomic" the difference between individual and equilibrium behavior. First, when $\beta = 1$, individual and equilibrium behavior are in fact identical. The differences arise solely out of the separation of private and social returns to labor. Second, the difference between individual and equilibrium behavior does not in any way depend on changes in the relative price of goods and leisure. (Relative prices are held fixed by the specification of the production technology.) The difference is due purely to an aggregate income effect.

Samuelson defines the fallacy of composition as: "A fallacy in which what is true of a part is, on that account alone, alleged to be also necessarily true of the whole" (1973, p. 14). In much of microeconomics, such reasoning is not fallacious because competition and individual optimization equate private and social tradeoffs. A primary claim of this paper is that the essence of macroeconomics lies in the difference between private and social tradeoffs.

The fallacy in the belief that the whole is merely the sum of the parts arises solely out of the difference between individual and aggregate marginal responses. If $\beta = 1$, then there is no fallacy in the "fallacy of composition"; individual and market labor supply respond identically. Looking at individual demands, equations (4), (5), or (6), we see that aggregate variables drop out when $\beta = 1$.

Contrast the result here with the more usual comparative static change in general equilibrium. In the usual case, a change in preferences changes demands, which may change prices, which may change demands further. Final demands, however, depend only on prices and individual endowments. (As a convenience, prices in my model are fixed by the production function.) Final demands in the model depend, in contrast, on prices, individual endowments, and the aggregate state of the economy.[4]

B. *The Failure of Say's Law*

In this section, I demonstrate that Say's Law fails. A unit increase in "supply" (endowment) fails to induce a unit increase in demand. Equilibrium output rises by less than the increase in endowments.

Suppose one unit of capital equipment rains down onto each worker. "Potential output" rises by the real rental rate, q. From equation (5), we see that individual demand for goods, k changing but Y and K assumed constant, rises by $\alpha_1 + \alpha_2$ times the real rental rate. From equation (6), we see that the increase in individual demand for goods plus

[3] It is not quite cricket to make welfare comparisons across regimes with differing utility functions. However, if we use the initial utility function to evaluate the new equilibrium consumption of goods and leisure, we see that utility has risen.

[4] A referee points out that in most general equilibrium models "...what is true for the individual is not true for the aggregate." The interesting part of the fallacy of composition is that individual behavior is changed by the aggregate state of the economy even after accounting for the equilibrium price vector. With a different choice of production technology, prices would change even when $\beta = 1$; so perhaps a more careful claim would be that when $\beta = 1$ there is no "interesting macroeconomic" fallacy of composition. Some readers may prefer the label "denial of aggregate income effect" to fallacy of composition.

leisure just absorbs the increase in potential output.

To see what actually happens in equilibrium, compare equation (7) to equation (5). Equation (7) shows that aggregate demand rises by βq, not q. The intuition behind the difference lies in the labor market. Greater endowments increase leisure demand. Individuals properly account for the private tradeoff involved. But individual compensation depends on aggregate L as well as individual l. Individuals properly regard aggregate L as exogenous. The increase in supply following the capital rain fails to produce enough demand to purchase that supply, as shown in equation (14). (Equation (14) is most easily derived from equations (4) and (9).) The failure of Say's Law is that demand rises by less than $(\alpha_1+\alpha_2)\cdot q$, not, of course, that demand rises by less than q.

$$(14)\quad \left.\frac{dGNP}{dK}\right|_{\text{aggregate}} = \frac{\beta}{1-(1-\beta)(\alpha_1+\alpha_2)}\cdot(\alpha_1+\alpha_2)q < (\alpha_1+\alpha_2)q = \left.\frac{dGNP}{dK}\right|_{\text{individual}}.$$

Notice that here again the failure arises solely from the separation of private and social response. If $\beta=1$, Say's Law holds; private demand and aggregate supply respond identically.[5]

The failure of Say's Law matters for its welfare consequences. Individuals naturally expect a one-unit capital rain to increase welfare by the return to capital. In equilibrium (see equation (11) and remember that F_l rises by q), net economic welfare rises by less than this amount. In the extreme case, where $\beta=0$, *NEW* rises not at all!

The failure of Say's Law is accompanied by, and due to, excessive underemployment. By construction, private marginal conditions hold; so there is no particular individual we can point to as being unable to obtain work at the going wage. But underemployment is a social problem. Every worker is selling less labor than she thought she would when the capital rained down.

III. Government Spending Policy

In this section I demonstrate two points about government spending policy. First, government spending can have Keynesian effects. It can increase *GNP* and reduce underemployment. Second, such a Keynesian program, that reduces individual choice, can increase individual welfare.

I consider the following type of balanced-budget government spending policy. The government (lump sum) taxes away consumption goods c from workers. With this revenue, it buys up government spending goods g from firms. Workers then consume the government goods purchased for them and allocate after-tax income to consumption and leisure.[6] I assume that the government does not directly redistribute income, in that taxes and government spending balance for each individual. Equation set (15) restates the individual optimization problem to include the government enforced choice of $\hat{g}$. "Hats" on variables distinguish this constrained problem from the earlier unconstrained choices.

$$(15)\quad \max_{\hat{c},\hat{l}} U=\alpha_1\ln(\hat{c}-\gamma_1)+\alpha_2\ln(\hat{g}-\gamma_2)+(1-\alpha_1-\alpha_2)\ln(l^*-\hat{l}-\gamma_3),$$

$$\text{subject to}\quad \hat{c}+\hat{t}=wE(\hat{l}|\hat{l}+\varepsilon)+qk,$$

$$\hat{g}=\hat{t}.$$

Please note that this really is a spending, not a transfer, policy. The spending policy forces a change in individual allocation between c and g.

[5]As with the fallacy of composition, the failure of Say's Law is interesting in that more happens than can be accounted for by price changes.

[6]The critical assumption is that the government can force individuals to consume more g than they would privately choose. In this model, any government spending which individuals can "undo" on private account is irrelevant.

Individual demands for constrained consumption and leisure are given by

$$(16)\quad \hat{f}_I = f_I - (\hat{g} - \gamma_2) + w(1-\beta)(\hat{L} - L);$$

$$\hat{c} - \gamma_1 = \frac{\alpha_1}{1-\alpha_2}\hat{f}_I;$$

$$l^* - \hat{l} - \gamma_3 = \frac{1-\alpha_1-\alpha_2}{1-\alpha_2}\frac{1}{w\beta}\hat{f}_I.$$

A little algebra will show that the constrained demands can be expressed in terms of the unconstrained demands, the constraint on purchases of g, and any change in aggregate labor due to the constraint on g:

$$(17)\quad \hat{c} - c = -\frac{\alpha_1}{1-\alpha_2}[(\hat{g} - g) - w(1-\beta)(\hat{L} - L)];$$

$$\hat{l} - l = \frac{1-\alpha_1-\alpha_2}{1-\alpha_2}\frac{1}{w\beta}[(\hat{g} - g) - w(1-\beta)(\hat{L} - L)].$$

General equilibrium is found, as before, by aggregating individual demand schedules. We find the constrained general equilibrium in

$$(18)\quad \hat{F}_I = F_I - (\hat{G} - \gamma_2);$$

$$\hat{C} - \gamma_1 = \frac{\alpha_1\beta}{1-(1-\beta)(\alpha_1+\alpha_2)-\beta\alpha_2}\hat{F}_I;$$

$$L^* - \hat{L} - \gamma_3 = -\frac{1-\alpha_1-\alpha_2}{1-(1-\beta)(\alpha_1+\alpha_2)-\beta\alpha_2}\frac{1}{w}\hat{F}_I.$$

Algebra again allows the constrained aggregate demands, in (18), to be rewritten in terms of the unconstrained demands, as in

$$(19)\quad \hat{C} - C = -\frac{\alpha_1\beta}{1-(1-\beta)(\alpha_1+\alpha_2)-\beta\alpha_2}(\hat{G} - G);$$

$$\hat{L} - L = \frac{1-\alpha_1-\alpha_2}{1-(1-\beta)(\alpha_1+\alpha_2)-\beta\alpha_2}\frac{1}{w}(\hat{G} - G).$$

From (19), constrained GNP and NEW can be written as a function of their unconstrained levels and of the extent of fiscal policy. (However, note the serious caveat below on the interpretation of NEW.)

$$(20)\quad \widehat{GNP} - GNP = \frac{1-\alpha_1-\alpha_2}{1-(1-\beta)(\alpha_1+\alpha_2)-\beta\alpha_2}(\hat{G} - G);$$

$$(21)\quad \widehat{NEW} - NEW = (1-\beta)\cdot\frac{1-\alpha_1-\alpha_2}{1-(1-\beta)(\alpha_1+\alpha_2)-\beta\alpha_2}(\hat{G} - G).$$

We are prepared now to consider the effect of government spending policy. Aggregate variables depend only on $\hat{G}$, mean endowments, and the common taste parameters because agents share linear demand functions with identical coefficients. However, some care is needed with regard to the redistributional aspects of government spending. Operating on the basis of equation set (17), we can define a government policy as specifying a constrained $\hat{g}$, relative to the freely chosen g, for each worker. Such a policy may require the government to have, in some sense, a great deal of information, as the required g level is likely to be different for each person.

A more easily "administered" policy is to require every worker to consume the same level of G. Take the mean level of G consumption from the policy of the preceding paragraph and make this the required level for every worker, $\hat{g} = \hat{G}$. (For the mean worker, the policies coincide.) While this policy has the advantage of simplicity, it may have severe redistributional consequences. Very poor workers, forced to spend all their limited income on G, are going to be unhappy. In addition, redistributional policies have the potential to change β, an interesting, but for our purposes unnecessary, complication.

The two policies have the same macroeconomic consequences. I will consider the former policy as its welfare consequences are more straightforward in that we arrange

matters to avoid any interpersonal comparisons of utility. In what follows, active fiscal policy consists of forcing each individual to increase consumption of good g, by an amount $\hat{g} - g = \hat{G} - G$. The absolute level of consumption of g varies across the population, but the amount of increased g consumption is the same for everyone.

A. *The Fiscal Policy Multiplier*

In this section, I consider several aspects of government spending policy. Our first task is to examine the impact on *GNP* and employment of an increase in government spending. Some care needs to be taken here. The mere fact that the balanced-budget multiplier is positive should not be taken to be significant. A positive multiplier only indicates that part of the welfare loss from constrained spending is taken up in lost leisure. The proper test of a multiplier is that the aggregate multiplier be larger than the multiplier evaluated for an individual holding constant aggregate variables. In other words, the multiplier is economically significant only if the increase in *GNP* is more than can be accounted for by private substitution away from leisure. The second task is calculation of the welfare consequences of increased government spending. A positive relation between spending and welfare is a further sign that we should take seriously a positive fiscal policy multiplier. Our third task is to find the optimal government policy.

Equation set (17) shows the effect on individual goods demand of a unit increase in G, holding constant aggregation variables. The *GNP* rises by $(1-\alpha_1-\alpha_2)/(1-\alpha_2)$, which lies strictly between zero and one. Call this effect the "individual multiplier."

The coefficient in (20) is the aggregate fiscal policy multiplier. The aggregate multiplier is greater than the individual multiplier. As β approaches one, the former approaches the latter. As β approaches zero, the aggregate multiplier goes to one, the simple Samuelson balanced-budget multiplier result. Note also that employment is increased by expansionary fiscal policy. So government spending affects the economy pretty much as predicted by a simple Keynesian model.

B. *Political and Welfare Implications of Increased Government Spending*

I use welfare analysis for two questions. First, will individuals favor increased government spending? Second, are they better off?

Does the individual want to pay taxes and receive government services in return? At first the answer would appear to be no, simply on the grounds that if the individual had wanted the services, she would have bought them in the first place. This reasoning is not quite correct. Because her individual full income is higher with higher government spending, the worker is not evaluating her utility function at the same margin as in the absence of government spending. We need to ask whether the worker feels she is receiving too much of good g holding constant the aggregate labor supply term in equation (16).

With or without government intervention, we know that the marginal rate of substitution between consumption and leisure equals the relative prices, $MU_L = w\beta MU_C$. Without government intervention, we also have $MU_G = MU_C$. With government intervention, the marginal utility of G declines. Since the fiscal policy multiplier is less than one, we know that consumption of C declines and that the marginal utility of C rises. Since underemployment declines, the marginal utility of leisure also rises. Therefore, with government intervention, the marginal utility of G is less than marginal utility of C, and analogously for leisure. We see then that every individual feels that government intervention is privately undesirable.

Does government spending improve social welfare? While equation set (18) shows equilibrium consumption and leisure only for the mean worker, a little algebra shows that equations (19) also apply on a person by person basis. In (22) I totally differentiate individual utility with respect to changes in government spending.

$$(22)\quad \frac{dU}{dG} = -MU_C\frac{\alpha_1\beta}{1-\alpha_1-\alpha_2+\beta\alpha_1} - MU_L\frac{1}{w}\frac{(1-\alpha_1-\alpha_2)}{1-\alpha_1-\alpha_2+\beta\alpha_1} + MU_G.$$

At very high levels of government spending, MU_G is very low and MU_C and MU_L become very high. So it is clear from (22) that too much government spending is a bad thing because people are forced to consume goods they don't want. Is any government spending beneficial? The relations between the marginal utilities make it easy to evaluate (22) at the unconstrained point in (23).

(23)
$$\left.\frac{dU}{dG}\right|_{\hat{G}=G}=MU_C\cdot\frac{(1-\beta)(1-\alpha_1-\alpha_2)}{(1-\alpha_1-\alpha_2)+\beta\alpha_1}.$$

So long as $\beta<1$, some government spending is beneficial. The greater the deviation between the private and social return to labor, the more valuable is increased government spending. In the extreme case, $\beta=0$, the first unit of government spending increases utility by the marginal utility of consumption, or in other words, the first unit of increased government spending is as valuable as a one-unit rain of income on each individual.

While utility is unobservable, net economic welfare is measurable in principle. Equation (21) shows *NEW* to be strictly increasing in $\hat{G}$, in apparent contradiction to the fact that utility first rises and then falls with increases in $\hat{G}$. The *NEW* values goods and leisure using relative market prices as a guide to marginal rates of substitution. When individuals are constrained to purchase more government services than they desire, price need not be proportional to marginal utility and *NEW* can be a misleading indicator of social welfare, although still better than *GNP*. Perhaps this problem can be best viewed as a version of the old observation that the practice of valuing government services at factor cost in the national income accounts leads in general to a mis-estimate of the value of total output.

C. *Optimal Fiscal Policy*

We find the optimal level of government spending by setting dU/dG in (22) equal to zero. Optimal government policy, shown in (24), is to increase spending above the unconstrained level. The greater the deviation between the private and social return to labor, the lower β, the more the government should spend relative to unconstrained G.

(24)
$$\hat{G}^*-\gamma_2=\left[\frac{1-\alpha_1-\alpha_2}{\beta}+(\alpha_1+\alpha_2)\right]\cdot[G-\gamma_2].$$

It is interesting to note that this policy involves setting G consumption for the mean worker at the first-best solution, though this is probably an artifact of choosing a utility function without cross-substitution effects.

To summarize this section, increased government spending increases *GNP* and reduces underemployment. A moderate amount of government spending causes a Pareto improvement in welfare.

IV. Extensions

I deal here with some possible extensions of the model and possible criticisms.

It is easily observed that in this model the government can force the economy to a first-best solution if it can subsidize labor at the rate $1/\beta$ and pay for the subsidy by levying lump sum taxes, as this would equate social and private marginal conditions. I expect that with sufficient heterogeneity, it would be impossible for the government to obtain a first-best solution without a great deal of information. Another way to say this is that in our economy it is difficult to see why a broad-based wage subsidy policy would be less intrusive than government spending programs.[7]

As it stands, this model does not look much like the aggregate demand models we are used to. If we wish, we can make the model take a somewhat more familiar form. Let us return to the constrained model. Aggregate demand is a function of aggregate

[7] In any event, Western economies tax, rather than subsidize, labor income. So the theoretical possibility of achieving a first-best result does not reduce the applicability of our results.

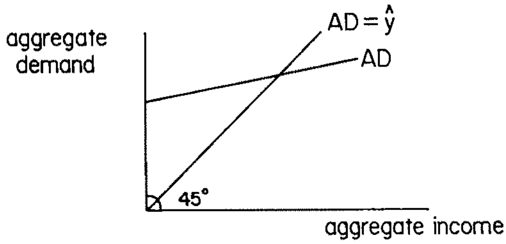


FIGURE 2

income and government spending, as in

$$(25)\quad AD = \text{constant} + \frac{1-\alpha_1-\alpha_2}{1-\alpha_2}\hat{G} + (1-\beta)\frac{\alpha_1}{1-\alpha_2}\hat{Y}.$$

Figure 2 shows aggregate demand graphed against the equilibrium condition that demand equals output. Samuelson's 45° line returns (with a little crowding out, since a unit step in G raises the aggregate demand line by less than a unit.)

A similar diagram can be drawn for the unconstrained aggregate demand system summarized by equation (7). One way this model differs from more familiar macro models is that here active fiscal policy changes all the multipliers as well as the level of *GNP*.

To reinforce the analogy with a Samuelson model, consider an increase in "autonomous consumption," caused by a unit increase in γ_1, in the unconstrained model. From equation (5) or from equations (4), we can see that this shift increases individual demand for goods by $1-\alpha_1-\alpha_2$. The marginal propensity to spend, by inspection from equation (7), is $(\alpha_1+\alpha_2)\cdot(1-\beta)$. For tradition's sake, let *MPC* represent the marginal propensity to spend. Application of a little algebra to equation (10) confirms that the increase in γ_1 increases equilibrium *GNP* by $(1-\alpha_1-\alpha_2)\times 1/(1-MPC)$.

There are several important aspects of aggregate demand missing in this model. According to the textbook pedagogy, there is a "leakage" from the "circular flow" out of income into saving. Most macroeconomic models are based on substitution between consumption and saving.[8] The model here is strictly atemporal. Substitution occurs between goods and leisure. The results look very "early Keynesian," omitting both what the profession has learned about life cycle/permanent income consumption and what we are now learning about intertemporal substitution of labor. As a consequence, we have no way to distinguish long-run from short-run behavior. The model is also "early Keynesian" in that every good is always in undersupply when compared to a first-best world. While utility can never be higher than in a first-best world, a different choice of the utility function might lead to an "oversupply" in a subset of markets.

Our model further diverges from more traditional views by pinning all its results on a labor market failure. Traditional models have focused more on goods markets failures than on labor market failures. I picked out this single deviation from complete Arrow-Debreu markets to demonstrate the general point that a rational competitive economy can exhibit Keynesian behavior. I have no argument with economists who suspect other failures are of greater empirical significance.[9]

Finally, this is a real markets model with no role for monetary aspects.[10] However, it is evident that money plays a great role in the economy. It is not clear how such a role would be integrated into this model or any of its near term successors.

[8] Much as the results in the preceding paragraphs look like a Samuelsonian model, there is no "paradox of thrift" in this model.

[9] In fact, Stanley Fischer, Mark Flannery, and Shelly Lundberg each independently suggested the same alternative to my labor market failure. Suppose, unlike my model, G is truly a public good, but that government spending can be financed only by a (distorting) income tax. Optimal government behavior will include some production of the public good. The taxation to support this production puts a wedge between private and social returns to work and might thereby induce the same possibilities for Keynesian fluctuation we have seen above. L. Jay Helms pointed out that a flat tax on labor income would produce analogous results with the regression coefficient β being replaced by one minus the tax rate.

[10] In this, I have followed Robert Solow's injunction to the profession to concentrate on real events rather than purely "monetary" theory.

V. Conclusion

A brief review of my premise and results is in order. The model I present is neoclassical, excepting only one well-specified market failure due to less than perfect information about individual labor productivity. In the absence of active fiscal policy, we see that Say's Law fails and I demonstrate why there is a "fallacy of composition." Both of these results follow from the fact that individual demands depend on aggregate income as well as prices and individual endowments. I show that at least some degree of expansionary fiscal policy is desirable, even though individual agents would rather be excused from participation. Stabilization policy is a public good.

This is the first, not the final, model on this topic. Its goals are therefore modest. It should however leave the reader convinced that competition and rational behavior are not inconsistent with observed Keynesian behavior.

REFERENCES

Lundberg, Shelly J. and Startz, Richard, "Private Discrimination and Social Intervention in Competitive Labor Markets," *American Economic Review*, June 1983, *73*, 340–47.

Nordhaus, William and Tobin, James, "Is Growth Obsolete?," in their *Economic Growth*, Vol. V, Ann Arbor: NBER University Microfilms, 1972.

Samuelson, Paul A., *Economics*, 9th ed., New York: McGraw-Hill, 1973.

______, "The Simple Mathematics of Income Determination," in L. A. Metzler et al., eds., *Income, Employment and Policy: Essays in Honor of Alvin Hansen*, New York: W. W. Norton, 1948, 133–55.

Unionization and Firm Performance: The Impact on Profits, Growth, and Productivity

By KIM B. CLARK*

The history of collective bargaining in the United States has been marked by dramatic episodes of confrontation that underscore the change unionization brings to the operation of an enterprise. Yet strikes and lockouts are only the most visible of pervasive differences in the employment relationship in organized establishments. It is well known that a wide variety of changes in the employment contract and adjustments in management procedure followed the wave of organizing begun in the 1930's.[1] In spite of the diffusion of many practices associated with collective bargaining (seniority, grievance systems) recent research has revealed the continuing existence of important differences between union and nonunion establishments in policies governing compensation, exit and entry, dispute resolution, and internal promotion.[2] As Richard Freeman (1980) has argued, these differences reflect the complexity of the employment relation and the potential for collective action to yield a different set of conditions in the presence of substantial information problems. Yet little is known about the effect of these differences on the profitability of the firm.

The large body of evidence on the union wage effect, for example, is not sufficient to establish a union effect on profits. Other changes in the employment contract may lead to firm or worker adjustments which either reinforce or offset the effect of increased wages on costs. The potential negative effects of work rules and protection of malfeasance are well known, but some recent evidence suggests that unionization may also lead to improvements in operations through reductions in turnover and changes in management procedures.[3] These considerations have motivated statistical comparisons of union wage and productivity differentials. Freeman and James Medoff have stated the assumptions underlying this approach quite clearly: "...Unionism may increase productivity in some settings and decrease it in others. If the increase in productivity is greater than the increase in average unit costs due to the union wage effect, then the profit rate will increase; if not, the rate of profit will fall" (1979, p. 81).

As I show below, this inference is only valid under quite restrictive assumptions. In general, it is not possible to infer changes in the rate of return on capital from information on union wage and productivity effects. The impact of the union on a firm's economic performance depends on the wage-setting process, the structure of markets, and the technology of production. Although an analysis of wages and productivity provides useful information about the operation of the firm under collective bargaining, an assessment of the impact of the union on profitability requires a direct examination. Furthermore, evaluating the efficiency consequences of unionization requires analysis of several measures of the firm's economic per-

*Graduate School of Business Administration, Harvard University, Soldiers Field, Boston, MA 02163 and National Bureau of Economic Research. This research has been supported by the Division of Research of the Harvard Business School and the Labor Studies program of the NBER. The Strategic Planning Institute provided access to the Profit Impact of Market Strategies (*PIMS*) data and supported the statistical analysis through a computer grant. I am indebted to David Ellwood, Therese Flaherty, Richard Freeman, David Garvin, Lawrence Summers, and participants in various NBER seminars for comments and suggestions. Phil Jarimyszin and Lorie Wilson provided excellent research assistance.

[1] The basic reference is Sumner Slichter, James Healy, and Robert Livernash (1960).

[2] A review of these findings is contained in Richard Freeman and James Medoff (1979).

[3] The papers by Freeman, by Charles Brown and Medoff (1978), and myself (1980) are representative.

formance. Unionization works through more than one mediating factor, and the impact of the union on a given measure of firm performance depends on the particular context in which bargaining and production take place. Thus, focus on a single indicator can be misleading.

This paper uses microeconomic data on over 900 product-line businesses to gauge the impact of the union on economic performance. In Section I, relatively simple models of the firm are used to derive a number of hypotheses about the effect of unionization. An important aspect of the analysis is the role of market structure and the institutional context of the wage-setting process. A clear implication of the theoretical analysis is the need to examine several indicators of firm behavior in order to draw inferences about the operation and consequences of collective bargaining. Section II presents an empirical analysis of unionization and interfirm differences in the rate of return on capital, sales growth, and productivity. Section III provides a brief summary and suggestions for further work.

I. A Theoretical Framework

The starting point for the models developed in this section is a single product monopolist with a constant elasticity demand curve and a constant returns production process. Although pure monopolies are rare, analysis of this case is useful because most of the firms to be dealt with in the empirical work face a downward-sloping demand curve and enjoy some barriers to entry. Unionization enters the analysis in two ways. I first treat the collective bargaining process as a problem of selecting a point on the firm's labor demand curve. In this context, the firm is assumed to be free to choose the level of employment and to adjust other decision variables in order to maximize profits. The second treatment of unionization allows the two parties to arrive at a wage-employment combination off the labor demand schedule. In this case, output and input decisions depend on the objectives of the parties and specification of the bargaining process.

The analysis yields results on several measures of firm behavior, but focuses particularly on the rate of return on capital as the basic measure of profitability. While a given firm's objective is to maximize total profits, some way must be found to scale total profits in order to provide a basis for comparison with other (possibly different-sized) enterprises. Because theory suggests that risk-adjusted returns should be equalized across industries and firms, the rate of return on capital has been widely used in empirical work. Other indicators of profitability, including various price-cost margins, have been suggested in the literature and will be examined in turn.

A. *The Monopoly Case*

The firm in this model is a profit-maximizing monopolist facing a demand curve with constant elasticity. For simplicity, the technology is specified to be *CES* with constant returns.[4] The supply of factors is perfectly elastic. Initially, I assume that the only effect

[4]The model developed in this section is a special case, but it illustrates the point that the impact of a union-induced wage increase on the rate of return on capital depends on the parameters of production and demand. Only in the case where the price of output is fixed does the union wage increase have an unambiguously negative effect on the rate of return on capital. This can be seen by expressing the rate of return on capital as

$$\pi/K=[PQ/L-WL/L]/(K/L),$$

and (with the price fixed) calculating the logarithmic differential. This yields

$$d\ln(\pi/K)=(1/(1-\alpha))[d\ln(Q/L)-(1-\alpha)d\ln(K/L)]-(\alpha/1-\alpha)d\ln W,$$

where α is labor's share, and W is the wage rate. Assuming that the changes in these variables were caused by unionization, the term in brackets is the union productivity effect, while the change in the wage is the union wage effect. With no productivity effect, and assuming that unionization raises the wage, the rate of return on capital will fall. If the price is allowed to vary (i.e., if the firm is assumed to have a downward-sloping demand curve), one obtains the same result as given in the text. In general, the direction of the effect also depends on the elasticity of demand and returns to scale.

of unionization is an increase in the wage. The firm takes the wage as given and chooses a level of employment consistent with its labor demand schedule. For the moment, let the union markup over the competitive wage be determined outside the model. The wage relationship can be written as

$$(1) \qquad w_u = w_0(1+mU),$$

where w_0 is the competitive wage, m is the percentage markup, and U is a dummy variable that has the value 1 if the firm is unionized. The effect of unionization on the firm can be developed from the solution of the firm's maximization problem, given by[5]

$$(2) \qquad \max_{K,L} \pi = f(K,L)^{(1-1/\eta)} - w_0 L - r_0 K.$$

The first-order conditions can be used to derive expressions for the optimal quantities of capital and labor. These, in turn, can be substituted into the definition of total cost to yield a cost function. Under constant returns (and perfectly elastic factor supplies) this function has the form

$$(3) \qquad C = A_0^{-1} g(w,r) Q,$$

where $g(w,r)$ is marginal cost, and A_0 is an index of total factor productivity. The optimal level of output is determined by the equality of marginal revenue and marginal cost. With a constant demand elasticity, marginal revenue is simply $(1-1/\eta)Q$ so that at the optimum (ignoring the constant A_0),

$$(4) \qquad Q^* = [g(w,r)/(1-1/\eta)]^{-\eta}.$$

Since the elasticity of marginal cost with respect to changes in the wage is simply α, labor's share, equation (4) implies that the elasticity of output with respect to wage changes is $\alpha\eta$.

Using (4) and the definition of profits, total profit can be written as

$$(5) \qquad \pi = \left(\frac{1}{1-1/\eta}\right)^{-\eta}\left(\frac{1}{\eta-1}\right)[g(w,r)]^{1-\eta}.$$

Thus, the elasticity of total profits with respect to changes in the wage is

$$(6) \qquad \eta_{\pi w} = \alpha(1-\eta).$$

These results provide clear predictions of the effect of unionization on output, the capital-labor ratio and total profits. Assuming $m>0$, unionization leads to a decline in output of ηm percent, a decline in total profits of $\alpha(1-\eta)m$ percent, and a percentage increase in the capital-labor ratio of σm, where σ is the elasticity of substitution.

The effect on the rate of return on capital is less clear-cut. Clearly, as long as $\eta>1$, unionization will lower total profits in the absence of offsetting productivity effects. If the stock of capital were unchanged, the rate of return on capital would also decline. But the capital stock will not remain constant. Whether it rises or falls depends on scale and substitution effects. In the case of a *CES* production function, this elasticity is

$$(7) \qquad \eta_{kw} = \alpha(\sigma-\eta).$$

Note that the scale effect is just the effect on output $(-\alpha\eta)$ because of the constant returns assumption.

The net effect of an increase in the wage on the return on capital can be derived by comparing (6) with (7) so that

$$(8) \qquad \eta_{(\pi/k)w} = \alpha(1-\eta) - \alpha(\sigma-\eta)$$
$$= \alpha(1-\sigma).$$

Thus, the rate of return on capital will fall with an increase in the wage if the elasticity

[5] Note that I have defined profits net of a required return to capital. Doing the analysis in terms of gross returns would change the magnitude of the calculated elasticities, but would have no effect on the signs; since the required rate of return is assumed to be fixed, the sign of changes in the gross rate of return (required plus excess) arising from a wage increase depends on the sign of changes in the excess rate of return.

of substitution is greater than one.[6] Otherwise, the return will rise.

How will these results change in the presence of a union productivity effect? Following previous work, the productivity effect can be characterized by a simple change in A, the index of total factor productivity. Let A_0 be the index before unionization, and d the union effect on productivity.[7] The index after unionization is then

$$(9) \qquad A_u = A_0(1+dU).$$

One way to illustrate the effects of unionization is to separate the effect of the productivity index from the effects of factor prices. Changes in total factor productivity have a direct effect on costs, while the wage effect depends on labor's share. The optimal level of output with both wage and productivity effects is

$$(10) \quad Q^* = \left[A_0(1+dU)\right]^{\eta} \times \left[g(w_0(1+mU), r)/1-1/\eta\right]^{-\eta}.$$

In elasticity terms, the effect of unionization on output is then

$$(11) \qquad \eta_{qu} = -\eta(\alpha m - d).$$

Thus, if the productivity effect is positive, and equal to αm, there will be no change in output under unionization. A negative productivity effect will simply reinforce the wage effect.

Equation (10) supports a similar conclusion about total profits. But the effect on the return on capital depends on relative changes in the capital stock. Using the notation developed in (10), the effect of unionization on total profits can be written as

$$(12) \qquad \eta_{\pi u} = (1-\eta)(\alpha m - d).$$

With a *CES*, constant returns production function, the expression for the effect on

[6] This conclusion must be modified if returns to scale are not constant. In the presence of nonconstant returns, the effect on output of changes in the wage depends on the elasticity of demand and the shape of the marginal cost schedule. Assuming constant elasticities, equation (6) becomes $\eta_{\pi w} = \alpha[1-(\eta/(1+\eta\varepsilon))]$, where ε is the elasticity of the marginal cost schedule. Letting θ indicate the returns to scale parameter, equation (7) can be written $\eta_{kw} = \alpha[\theta\sigma - \eta/(1+\eta\varepsilon)]$. Combining the two expressions, one obtains $\eta_{(\pi/k)w} = \alpha[1-\theta\sigma]$. With decreasing returns to scale, this elasticity is more likely to be larger for a given elasticity of substitution, and may be positive even if $\sigma > 1$. It remains true, however, that the effect of an increase in the wage on the rate of return depends only on the parameters of the production function.

[7] It is important to note that the effect of the union is assumed to apply to the production process as a whole, rather than to a particular factor of production. The assumption of a "neutral" productivity effect is consistent with existing evidence that unionization affects all aspects of the enterprise. (See my 1980 paper and Slichter et al.) If the union effect works through changes in the efficiency of labor alone, one must change the interpretation attached to the union coefficient in the productivity equations. Moreover, the theoretical argument about profitability must be modified as well. In order to illustrate the first point, consider a labor-only union effect in a Cobb-Douglas production function:

$$Q = AK^{1-\alpha}[L(1+cU)]^{\alpha},$$

where c is the union productivity effect on labor, and U is the union dummy. Taking logs, and rearranging terms gives

$$\ln[Q/c] \cong \ln A + 1 - \alpha \ln[K/L] + \alpha cU.$$

If the union effect were assumed to be neutral, it would appear as in equation (9) and the union term given in the productivity equation would be dU, instead of αcU. Estimation of the coefficient on the union dummy would not be affected, but interpretation of the coefficient would depend on the assumed route through which unionization affected costs.

On the theoretical side, a labor-only union productivity effect will affect the return on capital. The logic of this line of argument is that the productivity effect changes the capital-labor ratio, resulting in more capital than would be the case if the effect were neutral. Assume that the effect on total cost is the same in either case, and let the labor-only effect be indicated by d^*. Equation (12) is unchanged: $\eta_{\pi\mu} = (1-\eta)(\alpha m - d^*)$, while equation (13) becomes $\eta_{ku} = d^*\eta + \alpha m(\sigma - \eta)$. This differs from the neutral case, because there is no direct reduction in capital requirements per unit of output (it is assumed to be positive). Equation (14) then can be written as $\eta_{(\pi/k)u} = \alpha m[1-\sigma] - d^*$. With a labor-only productivity effect, the elasticity of the rate of return with respect to unionization differs from its value when only a wage effect is present. If $\sigma > 1$, the elasticity will be more negative; while with $\sigma < 1$, the elasticity will be less positive and may turn negative.

capital is

$$\eta_{ku} = d(\eta - 1) + \alpha m(\sigma - \eta). \tag{13}$$

Inspection of (12) and (13) reveals that the productivity effect enters symmetricallly, and thus has no effect on the rate of profit. Changes in the rate of profit depend on the elasticity of substitution:

$$\begin{aligned} \eta_{(\pi/k)u} &= (1-\eta)(\alpha m - d) \\ &\quad - [d(\eta - 1) + \alpha m(\sigma - \eta)] \\ &= \alpha m[1-\sigma]. \end{aligned} \tag{14}$$

As before, the rate of profit will rise or fall depending whether σ is greater than or less than 1. In this context, it is not possible to infer changes in the *rate of profits* from information on wage and productivity effects unless σ is known. Wage and productivity comparisons do yield clear inferences about total profits, output, and the capital-labor ratio.

B. *Alternative Bargaining Models*

The above analysis assumes that the wage-employment bargain lies on the firm's labor demand schedule. This assumption has significant analytical power, since it allows one to treat the problem of unionization using the theory of derived demand. However, two related aspects of the approach suggest the need for further analysis. First, a variety of bargaining issues associated with work rules, the introduction of technology and compensation seem to involve wage-employment combinations off the firm's labor demand schedule.[8] Second, in a bargaining context, points on the labor demand schedule are unlikely to be Pareto efficient. Unless the union has a fixed coefficient objective function, the contract curve will be off the demand curve. Since either the firm or the union could be made better off without reducing the welfare of the other party, it seems reasonable to expect that attempts would be made to move toward a more efficient bargain.[9]

Any attempt to achieve a bargain off the demand curve greatly complicates attempts to gauge the union's impact on firm performance. It is necessary not only to specify the technology and demand parameters, but the objective of the union and the specific bargaining process, including some statement of relative power must also be brought into play. A definitive treatment of those issues will not be attempted here. The more modest purpose of this section is to illustrate the implications of alternative bargaining setups using a relatively simple bargaining model.

For our purposes, the complexity of the union-management bargaining process can be simplified to the problem of choosing a point on a contract curve defined by the objective functions of the two parties. The assumption that the agreement lies on the contract curve implies that bargaining extends beyond wage rates to include, perhaps implicitly, the level of employment, and other aspects of production.[10] A variety of instruments are available that effectively constrain production adjustments by the firm without involving an explicit agreement on the number of workers, or the total hours of work. Work rules and provisions covering new technology can be interpreted as measures to achieve a bargain off the demand curve. Moreover, various compensations rules (royalty payments, profit sharing, equipment differentials) can have similar effects.[11]

[8]Frederick Warren-Boulton (1978) describes several such measures in his work on vertical control by unions.

[9]The likely inefficiency of points on the demand curve is a central feature of the work of Robert Hall and David Lilien (1979); for a recent treatment of the bargaining problem, see Ian McDonald and Robert Solow (1981).

[10]If the wage alone were subject to agreement, with the firm free to adjust output, prices, and other factors, the firm would choose a point on its demand curve. Note that the bargaining model implicitly assumes that the relationship between the union and the firm persists indefinitely. A richer analysis would deal with the question of maintenance of organization.

[11]See Warren-Boulton (pp. 119–56) for an intensive analysis.

Although in practice the outcome of bargaining where such provisions play a role will depend on the character and distribution of information, I will abstract from such difficulties here. I assume that sufficient information is available to allow the firm and the union to reach the contract curve. As above, the analysis concerns a single product monopoly operating under conditions of constant returns. The objective of the firm remains the maximization of profits. The union's objective is to maximize the difference between the wage bill and the opportunity cost of employment in the firm.[12] This can be written as $H=(w-w_0)L$, where w_0 is the wage in the absence of unionization, w is the realized wage, and L is labor input. The maximization problem confronting the firm and the union can be treated as a two-step process. In the first step, the joint interest or surplus of the two parties is defined and maximized by adjusting those variables that do not vary along the contract curve. The division of the surplus is determined in the second step through some process which reconciles competing claims.

Given the objectives specified above, the maximization problem can be written as

$$\max_{K,L} S = PQ - wL - rK + (w-w_0)L, \tag{15}$$

or

$$\max_{K,L} S = PQ - w_0L - rK. \tag{16}$$

The addition of the two objectives results in a maximand that is identical to the firm's objective in the nonunion setting. Under the bargaining regime, the firm makes production decisions as though it faced the nonunion wage. Thus, the stock of capital and the level of employment (and thus output and price) are unchanged after unionization. Only the wage varies along the contract curve. Given product demand and the opportunity cost of labor, the interests of the two parties are maximized at the point of maximum monopoly profits. The division of the surplus depends on the form of the objective functions, the distribution of power, and the particular decision rules which are assumed to apply. Although the total surplus is unchanged after unionization, workers share in profits, and thus reduce returns to the firm. With capital unchanged, the bargaining model unambiguously predicts a decline in the rate of return on capital.

The presence of union productivity effects in the bargaining context may lead to paradoxical results. A positive union effect, for example, results in an increase in output and total profits (before division with the union). Since the firm makes input and output decisions as though it faced nonunion wage rates, the union productivity effect has the same impact as would neutral technological change in the nonunion setting. The firm behaves as though marginal costs had declined and increases output (and lowers prices). Since the elasticity of demand exceeds one, the stock of capital and labor input are increased in proportion.[13] If the productivity effect is large enough, it may increase total profits sufficiently to leave the firm's profits unchanged after division with the union. Even in this case, however, the rate of return on capital will fall, since the stock of capital will have increased.

C. *The Implications of Market Structure*

The analysis thus far has been limited to the monopoly case. Differences in product markets are likely to influence the impact of the union in two ways. For any given wage or productivity effect, the response of the

[12] This assumption has been used by Sherwin Rosen (1969). The implication is that the union seeks to maximize the rents associated with bargaining collectively. This objective function is, of course, not the only objective one could imagine. Other functions lead to efficient bargains that influence output and the factors of production.

[13] A neutral change in technology (to which a union productivity effect is equivalent) has two effects on factors of production. The first is a direct reduction in input requirements equal (in percentage terms) to the productivity effect. The second is an increase in requirements through lower prices and increased output; the second effect depends on the elasticity of demand. If d is the productivity effect, the impact on K and L, for example, is $d(\eta-1)$.

firm to unionization may depend on the nature of competition it faces. Furthermore, the size of the wage and productivity effects themselves may depend on the structure of markets. This is particularly evident in the bargaining model, where relative power influences the division of total profits. The introduction of market structure considerations requires a richer theoretical framework than the monopoly model provides. Note that the effect of the union on profitability in that model does not depend on the elasticity of demand. Thus, little can be learned about market structure and the impact of the union on profits in the single firm framework where competition is defined solely in terms of the shape of a constant elasticity demand curve. The development of a full-scale model of competition in oligopolies is beyond the scope of this paper. It is possible, however, to suggest a number of important relationships between firm performance and unionization in different market settings using a fairly simple theoretical framework.

In the model developed here, the firm undergoing unionization (firm 1) produces a differentiated product, and has several competitors. Demands are assumed to be interdependent. In order to simplify the notation, it will be useful to group the firm's competitors into a composite, representative firm (firm 2). Technology is assumed to be constant returns, and may differ in the two firms. The reaction of firm 1 to unionization will depend on the degree of product differentiation and the output and pricing rules which govern competition. A convenient assumption is that the firms reach equilibrium by maximizing joint profits. This framework ignores interesting complications (for example, unions and rivalrous behavior), but is sufficiently rich to illustrate the basic issues.

Expressed in inverse form, firm 1's demand curve is

$$P_1 = f(q_1, q_2), \tag{17}$$

with the partial derivatives of (17) assumed to be negative. Firm 1's profits can be written as

$$\pi_1 = q_1[f(q_1, q_2) - a_1], \tag{18}$$

where a_1 is marginal cost, and q_1, q_2 are equilibrium values. With only a wage effect, unionization raises a_1 and the effect on π is given by[14]

$$\partial\pi_1/\partial a_1 = (\partial q_1/\partial a_1)[f(q_1, q_2) - a_1] + q_1[f_1 \partial q_1/\partial a_1 + f_2 \partial q_2/\partial a_1 - 1]. \tag{19}$$

The first term in (19) is the output effect, while the second term captures changes in the price-cost margin. It is the latter that is of central importance in determining the effect of unionization on the return on capital in different competitive situations. The effect of a wage increase on firm 1's capital stock will depend on the output effect and the elasticity of substitution. Since output affects profits and capital symmetrically under constant returns, output effects will cancel out. Changes in the return on capital will depend only on changes in the price-cost margin and the extent of substitution in production. However, variation in union impact with variations in market structure will depend solely on changes in the price-cost margin, since there is no reason to expect σ to vary with the structure of the product market.

As specified in (19), the change in the price of firm 1's product is the sum of two effects: a movement along its demand curve, and the reaction of competitors which causes a shift in demand. Whether the price-cost margin rises or falls in response to changes in costs is indeterminate. However, if the own-elasticity of demand is roughly constant, then $f_1\, \partial q_1/\partial a_1 \simeq 1$, and, movement along the firm's demand curve will leave the price-cost margin unchanged. Since $f_2 < 0$, and since competitors will react to a rise in a_1 by increasing q_2, firm 1's price-cost margin will fall. The implication is that an increase in costs in a competitive environment (i.e., high values of $f_2 \partial q_2/\partial a_1$) will put a

[14]The assumption of differentiated products is crucial here. Without it, joint profit maximization would imply that q_1 would go to zero, with some kind of *ex post* redistribution of profits. The analysis without differentiated products is equivalent to the problem of the multiplant monopolist.

squeeze on margins, more so than if the firm had substantial market power.

A further example suggests the plausibility of this result. Consider the case in which the newly unionized firm 1 is one of many small firms in an industry with a dominant firm. Suppose further that the small firms largely accept the dominant firm's price (products are only moderately differentiated) that is set in the usual way on the basis of a net demand function. A rise in a_1 in this setting (with the costs of other firms constant) is likely to lead to a declining price-cost margin for firm 1 since a relatively small firm will be unable to recoup cost increases by raising prices. In contrast, if firm 1 were the dominant firm, the price-cost margin would be likely to change very little.[15] As long as an identical union wage differential was imposed in the two settings, and as long as σ was identical, the effect on the return on capital would be more negative where firm 1 had less power.[16] While intuitively reasonable, the notion that unionization has a stronger impact on less powerful firms is only one among several theoretically possible outcomes.[17] Strong conclusions must await empirical evidence.

The effect of market structure on the union's impact is likely to be unaffected by productivity effects. As long as the union effect on productivity does not vary with market structure, and unless unionization leaves costs unchanged, qualitative conclusions about the effects of differences in competitive environments will be unchanged from the wage-only case. With unchanged costs, however, the profit effect depends only on the elasticity of substitution and is thus independent of market structure.

Conclusions about the effects of market structure are somewhat stronger if a bargaining model of wage determination is introduced. The analysis is simplified because a negotiated wage change in the bargaining model leads to no changes in output or prices in the unionized firm. While the profits of firm 1 are reduced, competitors are unaffected. From the competitors' standpoint, unionization of firm 1 only affects how the total profits of firm 1 (that are unchanged) are allocated. In this context the effect of different competitive environments is straightforward. Since unionization does not affect real variables, the issue is how post-union returns to the firm compare to the pre-union level of profitability. As long as the wage differential is the same in different competitive environments, the union effect on profits will depend on how pre-union profits per unit of labor vary with the competitive environment. Under the usual assumptions about firm behavior, this ratio will be higher and, consequently, the firm's share of total profits will be higher, the greater the market power of the firm. In effect, the more competitive firms must give up a larger fraction of lower total profits if the condition of constant wage differential is to be satisfied.[18]

A union productivity effect does not change the conclusion. If unionization affects productivity, the level and price of firm 1's

[15] This assumes that interdependence effects are minimal.

[16] It should be noted that the mental experiment here is to compare an otherwise identical firm in two different market settings. The issue is not the profitability of the unionized firm relative to its competitors, but rather the differences in profitability within the same firm in the two different environments.

[17] This conclusion can also be illustrated in a model where a large firm is limit pricing behind barriers to entry. If unionization leads to an increase in average costs and an increase in the limit price, there will be no observed effect on profits, as long as the new limit price is less than the profit-maximizing price. All of these examples are merely illustrative. Other models suggest different conclusions. For example, a Cournot model of duopoly implies that unionization of the smaller firm actually reduces margins by less than unionization of the larger firm. But this model abstracts from cross-price effects and assumes a common output price. I am indebted to Therese Flaherty for these examples.

[18] Note that the firm is assumed to face the same pre-union wage no matter what the product market environment is. With the same union differential it must be the case that post-union wages are identical. In a bargaining model, the post-union wage can be expressed as $w=((1-\theta)\pi^*)/L$, where π^* is total or pre-union profits, L is labor input, and θ is the firm's share of total profits. Note that θ also measures the extent of decline in the firm's profitability. Consider a firm in two environments, one a competitive situation (firm 1), the other a situation where the firm has much more market power (firm 2). Since post-union wages are equal, $((1-\theta_1)\pi_1^*)/L_1=((1-\theta_2)\pi_2^*)/L_2$. It is reasonable to expect $\pi_1^*<\pi_2^*$, and $L_1>L_2$. Thus, wages can only be equalized if $\theta_1<\theta_2$, which implies that profits must fall more in the competitive situation.

output will change, and adjustments will be made among competitors in order to maximize joint profits. Consider the case of a positive union effect, accompanied by a wage increase, both of which are independent of market structure. Consistent with the two-stage bargaining procedure developed above, the firm is assumed to incorporate the productivity effect into its calculation of total profits, before the surplus is divided. Thus, the productivity change affects the calculus of joint profit maximization in a manner exactly analogous to equation (19). As before, assessment of the difference in union impact in different competitive settings depends on changes in the price-cost margin. Let the relationship between total profits before and after the productivity effect be given by $\pi_1^* = (1+\delta_1)\pi_1^0$, where δ_1 captures the cross-price effects of competitors (the own-price effect is assumed to just offset the cost change) which will be positive. The firm's share of total post-union profits is $\pi_1/\pi_1^* = \theta_1$, and the union effect on profits is given by $\pi_1/\pi_1^0 = \theta_1(1+\delta_1)$. (Similar expressions can be written for the firm in the less competitive setting, firm 2.) As in the wage effect only case, the constant wage differential implies that $((1-\theta_1)\pi_1^*)/L_1 = ((1-\theta_2)\pi_2^*)/L_2$. Substituting for π_1^*, and recalling that $\pi_1^*/L_1 < \pi_2^*/L_2$, the equality of wages implies that $\theta_1(1+\delta_1) < \theta_2(1+\delta_2)$. Thus, the union has a smaller impact in the less competitive environment.[19]

D. *Market Structure and Wage/Productivity Effects*

The simple model of oligopoly developed above suggests that competition increases the response of firm performance to unionization. This result rests heavily on the assumption that wage and productivity effects are independent of market structure. Yet it has long been argued that wage determination is affected by structural conditions in the product market.[20] Moreover, the nature of competition is likely to affect both the opportunity and the incentive for the firm to improve productivity after unionization. In both cases, however, the direction of the effect is ambiguous. Although definitive predictions cannot be derived, theoretical considerations together with existing empirical evidence can be used to provide insight into likely directions of influence.

In the case of compensation, two conflicting effects govern the relationship between market structure and the union wage effect. Firms with substantial market power tend to earn above-average returns, which become the target of union negotiations. From the union's perspective, high profits increase the firm's ability to pay and constitute a legitimate ground for high wage demands. Furthermore, if returns substantially exceed the opportunity cost of capital, a high wage level need have little effect on the firm's long-term survival. In contrast, unionization of a firm operating in a highly price-competitive market may lead to very low long-term wage gains. Either the union will recognize the need to moderate demands, or the firm will not survive. Thus the only competitive firms observed are those with wages close to nonunion levels.

The foregoing argument ignores the role of relative power in wage determination. While it is true that, in the extreme case of perfect competition, the scope for wage gains is constrained, within the set of firms exercising some market power, the union wage differential may be affected by the relative power of the employer. Because of its ability to weather a strike, a firm with few competitors and substantial financial resources may be less willing to make concessions than a firm with less market power. Thus, unionization of a firm with high market share may not lead to a larger differential over nonunion wages. This is particularly true if the high market

[19] The discussion here abstracts from any rivalrous behavior. It is assumed that competitors accept reformulation of prices and outputs caused by the union. But it may be that competitors come out worse off, since some of the union's gains may come out of their profits. Whether the consensus holds up or breaks down may depend on the relative power of the unionized firm. Thus, even in the case of rivalrous behavior, it is likely that unionization will have less impact on the more powerful firm.

[20] See the papers by Leonard Weiss (1966) and Martin Segal (1964) for a discussion of the theoretical and empirical issues.

share firm has already foregone some profits by paying relatively high wages. The desire to increase an applicant queue to avoid community censure or unionization has been suggested in the literature as an explanation of a positive correlation between market share or concentration and the level of wages. In such a setting, the added impact of unionization may be small.[21]

Existing empirical evidence on the relationship between market structure and the union wage differential in U.S. manufacturing suggests that the constraint of competition and the power of the firm tend to be offsetting. Work by Leonard Weiss and by Freeman and Medoff (1981) suggests no statistically significant relationship between concentration and the union wage effect.[22] While the direction of the effect is generally negative (i.e., the differential is smaller in concentrated industries), large standard errors preclude strong inferences. The implication is that the wage effect is unlikely to offset, and may reinforce, the tendency for the union impact on profits to be more negative as the number of competitors increases and as market share declines.

Whether this result continues to hold in the face of productivity effects, however, is unclear. If positive union productivity effects are larger and/or more likely in more competitive settings, the union impact on profitability (and output) could be attenuated. The threat of economic demise may be more apparent the more competitive the market, and thus could provide compelling pressure for the firm to improve operations. In the case of perfect competition, for example, any union wage effect would appear to require some offsetting productivity gain to ensure survival. Yet pressure to improve operations is usually presumed to be present in highly competitive markets without unionization, leaving no possibility of further improvements after the union is introduced. If, as before, attention is confined to the set of firms exercising some market power, then the potential for improvement would seem to be greater in less competitive markets since organizational slack, or x-inefficiency is likely to increase with market share.[23] Whether potential gains are realized, however, depends on the presence of pressure from sources besides other competitors. Even though market-based pressure may be less, for example, the internal pressure imposed by the union could motivate a search for new methods and procedures which improve operating performance and offset, perhaps partially, any union wage gains and inefficient work rules. Furthermore, a vigorous organizational response to unionization could be observed in high market share firms if satisficing behavior is important in less competitive markets, and if the union threatens to reduce firm performance below satisfactory levels.[24]

In light of the empirical evidence on the union wage effect and market structure, the preceding argument suggests that the union impact on profitability may well be greater in firms with less market power, even if all firms experience some positive union productivity effects. If the more powerful firms are better able to recoup wage increases by altering internal organization, the tendency for competition to sharpen the impact of the union will be reinforced.

The theoretical analysis in this section suggests that the impact of the union on various indicators of firm performance depends on technology, the wage-setting process, and

[21] This argument is made by Weiss, who also finds some support for the queue explanation. Similar conclusions apply to the work of Orley Ashenfelter and George Johnson (1972).

[22] Additional evidence can be found in the paper by Thomas Pugel (1980).

[23] It is evident that such considerations apply only to firms enjoying some barriers to entry. The tendency for inefficiency to increase with market share has been inferred from the quadratic nature of the relationships between rates of return and market share. See especially the work by William Shepherd (1972).

[24] The analysis has not treated the potential impact of unionization on market structure. While principal concern is with the effect of the union on the firm, it is clear that the extent of unionization within an industry could affect the nature of competition, both by affecting entry conditions and communication about costs. Furthermore, if the union pursues a common wage policy, marginal competitors could be eliminated. The net result of these considerations is likely to be a larger impact of the union on less powerful firms.

TABLE 1—UNIONIZATION AND FIRM PERFORMANCE: THEORETICAL PREDICTIONS

Indicators of Firm Performance	Bargaining Regime[a] On the Demand Curve (1)	Bargaining Regime[a] On the Demand Curve (2)	Bargaining Regime[a] On the Contract Curve (1)	Bargaining Regime[a] On the Contract Curve (2)
Q/L	0	+	0	+
Q	–	0	0	+
π/K	[b]	[c]	–	–

Note: Cols. (1) = Wage Effect only; Cols. (2) = Wage and Offsetting Productivity Effect.

[a] Except where noted, the entries refer to the single firm model. In general it is the magnitude and not the signs which depend on market structure. Note also that the signs on Q/L assume that the capital-labor ratio is held constant.

[b] In the monopoly model, the sign depends on σ; with competition, the sign is more likely to be negative.

[c] The sign depends on σ.

market structure. The basic qualitative results are summarized in Table 1. It is clear that inferences about the net efficiency consequences of the union based on a single indicator such as productivity or profitability may be misleading. Evidence about productivity is incomplete without information on wage differentials and, even then, any resultant changes in the rate of return on capital may convey information about the extent of substitution possibilities, but little else. Similar conclusions apply to other indicators taken in isolation. A finding, for example, that unionization has no effect on growth could result from a bargaining model of wage determination, and need not imply the existence of a productivity dividend. Adding information on productivity (and other indicators) is thus essential. Only evidence on the pattern of effects over a number of indicators yields insight into the consequences of unionization.

II. Empirical Analysis

The analysis of Section I focused on three indicators of firm performance, each of which may be affected by unionization. The approach in the theoretical analysis has been to treat the two effects of unionization—wages and productivity—in the context of a reduced-form equation for firm performance. In the case of profits, for example, the effect of the union is gauged through the following expression:

$$\pi = B\left[A_0 e^{\lambda T}(1+dU)\right]^{\eta} \times \left[g(w_0(1+mU), r)\right]^{1-\eta}. \tag{20}$$

Clearly, profits and other measures of performance depend on technology, exogenous factor prices, the structure of markets, and the nature of demand. Unionization is assumed to affect performance through changes in wages and through its effect on productivity. While the various dimensions of performance are clearly related to one another, the intent is not to estimate a structural model of performance, but to identify the impact of unionization in otherwise comparable enterprises. The general approach, therefore, is to estimate a set of reduced-form equations which relate indicators of performance to a similar set of control variables (including unionization) which are treated as exogenous or predetermined.

The analysis makes use of micro data on individual "business units"—generally divisions of a given corporation—which are referred to as the "firm." The indicators of firm performance are defined as follows:

Rate of return on capital (ROI): net income (pretax) divided by total capital invested in the business;

Output growth (GS): rate of growth of deflated sales (price deflator is the firm's estimate of an index of the price of output);

Labor productivity (QL): sales (or value-added) divided by total employment.

These indicators reflect in part the theoretical analysis of Section I, in part the availability of data. Profitability is measured as the return on total capital invested, a measure long used in economic analysis and in business as a measure of performance.[25] "Total

[25] The net income measure is based on conventional accounting practices which treat $R\&D$, advertising and other marketing costs as expenses. Neither income nor

capital invested" has been used, since there is no meaningful distinction between equity and debt capital in the data (a line of business). Although primary emphasis is placed on the *ROI* measure, the impact of unionization on the return on sales (*ROS*—a price-cost margin) will also be examined. A second measure of profitability may be useful in interpreting the source of union-nonunion differences in the return on capital (for example, lower gross margin vs. higher capital intensity).

The effects of unionization on output and growth are captured using the rate of growth of sales deflated by an index of product prices. Although the theoretical analysis dealt with the effect of unionization on the level of output, the empirical work focuses on the question of whether unionization affects the extent to which a firm expands or contracts relative to its market. The only variable close to the theoretical output measure is the firm's market share. While results with market share will be discussed, this variable may be less a measure of performance (at least current performance), and more a measure of structural conditions in the market.

The productivity analysis uses two measures, sales per employee and value-added per employee. While it would be desirable to have better measures of labor input (hours worked, more detailed categories of employment), they are not available. In addition, the only price information available is the firm's estimate of its own selling price expressed as an index, with $1973 = 1.00$. Without cross-firm variation in the level of prices, only sector-specific price deflation (for example, 2-digit SIC) will be possible. As discussed below, an attempt will be made to control for variation in the ratio of materials to labor and the capital-labor ratio, and thus to estimate the effect of unionization on total factor productivity.

The equations to be estimated can be expressed in general terms as follows:

$$(21)\quad ROI_{it} = ROI(UN_i, \overline{MF}_{it}, \overline{MI}_{it}, \overline{H}_{it}, \overline{T}_{it}),$$

$$(22)\quad GS_{it} = GS(UN_i, \overline{MF}_{it}, \overline{MI}_{it}, \overline{H}_{it}, \overline{T}_{it}),$$

$$(23)\quad QL_{it} = QL(UN_i, \overline{MF}_{it}, \overline{MI}_{it}, \overline{H}_{it}, \overline{T}_{it}, KL_{it}, ML_{it}),$$

where for the ith firm at time t, UN_i is a measure of unionization, $\overline{MF}_{it}$ is a vector of variables measuring the nature of the competitive environment in the firm's narrowly defined market, $\overline{MI}_{it}$ captures the market structure of the firm's larger industry, $\overline{H}_{it}$ is a vector of work force characteristics, $\overline{T}_{it}$, is a technology vector, KL_{it} is the firm's capital-labor ratio, and ML_{it} is the materials-labor ratio. (Note that the value-added version of (23) omits the materials-labor ratio.) The specification assumes that the price of capital is constant across firms; the inclusion of work force characteristics is intended to control for variation in competitive wages, and in labor quality. A full listing of the variables and definitions included in the analysis is presented in Table 2.

For the most part, the models spelled out in equations (21)–(23) are straightforward. Performance depends on structural conditions, both outside the firm in its labor and product markets, as well as inside, in its technology. In practice, there is little information about technology and differences across firms in this dimension are captured by 2-digit SIC industry dummies and time trends. The data provide a richer set of control variables for market structure. Indeed, equations (21)–(23) depart from typical industry-level analysis in their specification of the firm's market structure, in the richness of the control variables, and in the dependence of productivity on market structure. In addition, the reduced-form character of the analysis results in the presence of work force characteristics (albeit industry-level char-

capital have been corrected for inflation in the results reported below. Capital is valued at historical costs and net income is taken as reported. However, a relatively crude attempt was made to restate capital in current prices, and to correct for profits on inventories. (The *PIMS* data provide an estimate of the replacement value of plant and equipment.) Since this correction was a very rough cut, and since it had no effect on the estimated impact of the union, results with the conventional profitability data are presented below.

TABLE 2—CONTROL VARIABLES

Variable	Definition	Source
Firm Control Variables (*KL*, *ML*, *MF*)		
LKL	Log of total investment per employee; total investment includes net plant and equipment and working capital	*PIMS*
LML	Log of purchased materials per employee	*PIMS*
MS	Share of served market (in percent)	*PIMS*
BIG3	Market share of three largest competitors (in percent)	*PIMS*
COMP	Number of competitors	*PIMS*
CPUR	Percentage of immediate customers that account for 50 percent of business' total sales	*PIMS*
EXPER	Approximate age of business (years since first commercial sale)	*PIMS*
RMG	Rate of growth of real market sales; (total market sales deflated by estimated market prices)	*PIMS*
Industry Market Structure (*MI*)		
INDG	Average rate of growth of sales in business' 4-digit SIC industry over 10-year period preceding business' entry into sample	*PIMS*
INDC4	Four firm concentration ratio in business' 4-digit SIC industry	*PIMS*
INDIMP	Share of imports in sales in business' 4-digit SIC industry	*PIMS*
COST	Cost disadvantage ratio of 3-digit SIC industry	*1972 Census of Manufacturers*
Industry Labor Market (*H*)		
RGN (1,2,3)	Fraction of shipments in Northeast, South and West for the firm's 3-digit SIC industry	*1972 Census of Manufacturers*
TEN	Average years of tenure in the firm's 3-digit SIC industry—1973	*CPS*, January 1973
GRD	Mean grade attended by employees in the firm's 3-digit SIC industry—1970	*1970 Census of Population*
AGE	Mean age of employees in the firm's 3-digit SIC industry—1970	*1970 Census of Population*
BLK	Percent of nonwhite employees in the firm's 3-digit SIC industry	*1970 Census of Population*
FEM	Percent of female employees in the firm's 3-digit SIC industry	*1970 Census of Population*

acteristics, tenure, years of education, demographics) in equations explaining profitability and growth.

The principal concern of the analysis is estimation of the overall impact of unionization, and its impact in different competitive situations. Industry-level measures of market structure include barriers to entry, industry growth, concentration, and the share of imports. The analysis assumes, however, that conditions in the firm's own narrowly defined market are the primary structural determinants of competition; this is the focus of the analysis of unionization and market structure.

Three variables are used to characterize the competitive environment the firm faces: the market share of the three largest competitors, the number of competitors, and the firm's own market share. These variables are intended to capture the size distribution of competitors, the difficulty of anticipating competitor reaction, and the firm's relative position. A variety of alternative measures of market structure using combinations and permutations of these variables (for example,

marginal concentration ratios, the Herfindahl index) were examined in the course of the empirical work, but were found to provide no additional insight or explanatory power.

The market structure variables enter the equations for profitability and productivity, but the firm's market share has been excluded from the growth equation to avoid redundancy. In the productivity equation, market structure is included to capture any x-efficiency effects or effects which may flow from relative size.

The unionization variable is based on the firm's response to the question: "Of all the employees in this business, what percentage are unionized?" The questionnaire makes no attempt to define what "unionized" means. Moreover, given the breadth of the work force involved ("all employees"), there is likely to be a good bit of error in the estimates. Fine distinctions between firms are not likely to be meaningful. It is much more realistic to assume that the respondents will accurately identify whether a business is completely nonunion, or whether the fraction unionized falls in a broad range (say, 30–60 percent). Most of the analysis, therefore, will use dummy variables to capture the union-nonunion distinction, rather than relying solely on variation in the percentage unionized.

Before examining the source of the data in more detail, a brief word about the exogenous variables in equations (21)–(23) may be useful. As noted at the beginning of this section, I treat unionization, market structure, and other control variables on the right-hand side of (21)–(23) either as exogenous or predetermined. A few of the variables are affected by the firm's decisions and their use raises the issue of simultaneity bias in the coefficients. Two variables deserve particular mention: the firm's market share and unionization.[26]

The market share of the firm provides information about the elasticity of demand and the firm's position relative to its competitors, characteristics of the market that are likely to have a major impact on profitability and that may affect productivity. I assume that these are structural features of the competitive environment, and are not affected by current profitability or productivity (the two equations where market share enters). It is clear that performance will affect structure over time, but existing evidence suggests that the lags are likely to be quite lengthy.[27] Of course, this assumption may be more tenuous for rapidly growing markets, particularly where products and firms are not well established. In order to test the sensitivity of the conclusions about union impact to the inclusion of market share, equations where market share is excluded will be estimated.

A further critical assumption in model specification is the exogeneity of unionization. At issue is whether estimated coefficients on the union variables capture the impact of unionization, or whether they also reflect the process through which establishments are organized (for example, unions organize the least, or most, profitable firms). The view that unionization is exogenous finds some justification on empirical grounds. For the great majority of unionized firms in this sample, union status was achieved prior to the time period covered by this data. Direct simultaneity is thus unlikely. It is, of course, possible that unionization could be correlated with an unobserved "firm effect" that persists through time, thus creating a correlation between union status and the error terms in equations (21)–(23). Such an argument is most plausible if one interprets the firm effect to be a "management quality effect."[28] The data set I use here is not suitable for study-

[26]The capital-labor ratio is also endogenous, but it is used to control for the effect of the union on productivity which works through wages. The union coefficient is thus to be interpreted as an estimate of d, the union effect on total factor productivity.

[27]See the paper by Stephen Martin (1979) for estimates of the lagged relationship between performance and structure.

[28]Other kinds of firms characteristics that might affect profitability, such as specialized resources, brand names, and so forth, are difficult to link to the process of unionization.

ing the firm effects and unionization, because no firm changes union status. Yet there are reasons to question the empirical relevance of the theoretical connection between a persistent "management quality" effect and unionization.

In the first place, evidence on the process of unionization from the 1940's and 1950's (the period when most of the firms in the sample were unionized) suggests that organization occurred in firms with a range of profit and management experience.[29] It was not uncommon for entire industries, or at least a sizeable majority of firms within an industry, to be organized. The resulting pattern of organization thus covered the gamut of performance and managerial skill. At the level of the individual establishment, the variety of actors involved in the organizing process argues against a simple relationship between management quality and organization. Union organizers, for example, may have found high profits (caused by high quality management) an attractive target for wage demands and thus may have devoted resources to organizing such firms. In contrast, poor management may be associated with dissatisfaction in the work force, and thus demands for union services. There is also substantial evidence that, in specific instances, union organization reflected forces unrelated to management quality, including community values and attitudes, and personality conflicts within the workplace.[30]

The historical evidence thus suggests an ambiguous connection between a firm effect associated with management quality and unionization. In addition, the requirement that the effect persists through time is problematic. Beyond the fact that competition is likely to erode, if not completely eliminate, any management quality effect on profitability, it is likely that unionization itself will affect the organization, personnel, and practices that existed prior to unionization and that underlie the firm effect. Whether the union's effect on the firm is positive or negative is an open question, but there is little doubt that unionization *changes* the management of the enterprise. These changes involve transfer and hire of personnel, changes in organization structure, operating rules and procedures, and management style.[31] There is, therefore, some basis for expecting a change in any correlation between unionization and the firm's quality of management.

The empirical work in this paper treats unionization as an exogenous variable and thus assumes no correlation between the union variable and any firm effects. Although there are a variety of estimation strategies for dealing with the possible unionization-firm effect connection (for example, instrumental variables, mixed logit models), the approach that is most appealing is to analyze the performance of the firm before and after unionization.[32] As noted above, however, there are no changes in union status in the *PIMS* data set, so that the before-after analysis is not possible. While tests of the assumed exogeneity of unionization will require a different set of data, I can indicate how inferences about the effect of union-

[29] The literature on the growth of unionization in the period 1935–1955 is extensive. See, for example, the work by Irving Berstein (1969) and Walter Galenson (1958).

[30] My research (1983) on management response to unionization underscores the importance of personal relationships and community attitudes in union organizing. See the work of Joel Seidman et al. (1951, 1958) for additional evidence on the organization of unions.

[31] The basic source on management change in response to unionization is Slichter et al. See also my earlier papers.

[32] The mixed logit models in which unionization is treated as a dichotomous endogenous variable depend on functional form for identification, while the instrumental variables procedure depends on finding variables that are correlated with unionization but uncorrelated with the firm effect, and not already included in the profit regression. The dependence on an assumed functional form for identification is not very satisfying from a substantive viewpoint. And in the case of instrumental techniques, the selection of instruments for unionization is likely to wind up resting on a set of assumptions that are no less problematic than the assumption that unionization is uncorrelated with firm effects. The best procedure would thus seem to be the before-after analysis (which amounts to estimation with firm dummies) referred to in the text.

ization in this sample may be affected by endogeneity.

A. *The Data Set*

The data used in the empirical work provide fairly detailed information on performance, technology, market characteristics and union status. Data are available on over 900 product-line businesses which participate in the *PIMS* project directed by the Strategic Planning Institute.[33] The Institute is composed of over 250 member companies which participate in the project by filling out annual statistical questionnaires on individual businesses within the company. Data are currently available from 1970–80. Not all businesses provide data for each year so that the design of the data base is unbalanced. In addition to income statements and balance sheets, the data set includes variables measuring the structure of the market in which the firm participates, and the competitive strategy of the business.

The data have a number of features which may affect interpretation of the results. The companies in the project tend to be large, diversified corporations; many are found in the *Fortune* 500, and almost all of them are found in the *Fortune* 1000. The analysis thus deals with the effects of unionization among a set of firms which are not representative of all firms in a given sector, but which account for a significant fraction of the assets and people employed. Because the unit of observation is a line of business, a given company may report on several businesses. The businesses tend to be synonymous with operating divisions of the company (for example, the washer and dryer division of the Diversified Appliance Corporation), but may include product lines within the division (commercial washers and dryers) if the appropriate data are available, and if the finer breakdown is useful to the company. In order to ensure consistency across firms, the definition of a business is determined by the company under specific guidelines established by the Institute.[34] Each business is identified by its 4-digit SIC industry. The unit of observation should be viewed as a "line of business" subcomponent of each 4-digit industry.

The central variables (for example, profits, capital stock, sales, etc.) used in this study are generally taken from the firm's income statements, balance sheets, and personnel records. Variables measuring the structure of the firm's 3- or 4-digit industry have been taken from the *Census of Manufacturers*, or the *Census of Population*, as appropriate. In some instances the firm is required to provide estimates of activity outside the firm (for example, the size of the firm's market, the number of competitors, etc.). As with business definitions, these estimates are processed under the Institute's guidelines, but the sources of information may differ across firms. In addition to these measures, the data set includes several variables that are based in part on perceptions of the respondents (for example, the relative quality of the firm's products). None of these variables, however, has been used in estimating the effect of unionization on firm performance.

Like all survey-based data sets, the *PIMS* data set is likely to be subject to reporting errors. But two considerations suggest that the data are of relatively high quality. First, the information requested is information of great value to the firm (for example, its

[33] The *PIMS* data have been widely used in studies of firm performance. See, for example, my paper with Zvi Griliches (1982), David Ravenscraft and F. M. Scherer (1981), and Lynn Phillips et al. (1983).

[34] The guidelines identify an individual business as a unit of the firm that 1) sells a distinct and closely related set of products; 2) to an identifiable group of customers; 3) in competition with a well-defined set of competitors. An important criteria in defining a business is the ability of the manager of the business both to design and execute a single strategic plan covering all phases of activity—marketing, manufacturing, research and development, and so forth. This requirement implies a measure of homogeneity in the products, customers, and competitors included in the definition of the business. In addition to establishing guidelines for defining businesses, the Institute assigns a "service coordinator" to each firm. The service coordinator works with member companies to ensure that the definitions used fit the guidelines; not only for defining businesses, but for selecting and measuring other variables as well. For more information on the procedures used see the *PIMS Data Manual*.

TABLE 3—MEANS AND STANDARD DEVIATIONS (*SD*): TOTAL, UNION AND NONUNION SAMPLES, 1970–80

	Total ($N=4681$)		Union ($N=3246$)		Nonunion ($N=1435$)	
Variable	Mean	*SD*	Mean	*SD*	Mean	*SD*
ROI	21.5	25.6	20.6	24.4	23.5	28.1
ROS	9.0	11.5	8.7	10.6	9.5	13.4
GS	9.2	28.8	8.1	28.0	11.6	30.3
LVL	3.28	0.57	3.28	0.54	3.27	0.65
LSL	3.89	0.61	3.90	0.58	3.85	0.66
LKL	7.78	0.71	7.80	0.68	7.75	0.79
LML	7.57	0.85	7.61	0.82	7.48	0.90
%UN	41.6	32.8	60.0	21.2	–	–
MS	22.6	18.8	22.6	18.2	22.6	20.0
BIG3	46.9	18.5	47.1	18.2	46.5	19.2
COMP	13.9	13.2	13.9	12.9	14.0	14.0
CPUR	13.5	11.2	13.3	10.9	13.9	11.7
EXPER	25.8	15.4	28.6	15.0	19.6	14.6
RMG	4.9	21.8	4.3	21.4	6.5	22.5
INDG	8.2	4.9	7.9	4.5	8.9	5.6
INDC4	47.5	24.2	48.1	23.5	46.2	25.6
INDIMP	4.7	6.1	4.9	6.4	4.2	5.3
COST	0.12	0.20	0.12	0.21	0.10	0.18

Source: Calculated from the *PIMS* data set with additions.

market share) and it seems reasonable to suppose that the firm is both in a position to know, and has expended effort to acquire accurate data. Second, a firm's participation in the project is motivated by a desire to use its data in the models developed by the Institute. Considerable effort is made to preserve confidentiality and ensure quality: participating firms only have access to their own data; sensitive data (profits) are reported in disguised or ratio form (for example, return on sales); researchers at the Institute run the data through an elaborate procedure to check for consistency and follow up gross errors with the company.

B. *The Evidence*

Table 3 presents means and standard deviations for selected variables. The calculations are broken down by union and nonunion status, and are based on data for North American (predominantly U.S.) manufacturing businesses over the period 1970–80. Each firm-year is treated as a separate observation, though some firms are not represented in each year. There are 902 firms in the sample, and thus an average of about 5 observations per firm (there are 4,681 observations in total).[35]

It is evident from the calculations that the unionized firms are substantially older, and participate in markets that are growing somewhat less rapidly than their nonunion counterparts. They tend to confront about the same number of competitors, and to compete in markets where the largest firms have a slightly larger share of the market.

As far as performance is concerned, the raw means suggest that nonunion firms are more profitable and that their sales are growing more rapidly, even more than could be accounted for by faster market growth. In terms of market share, however, there is no

[35]The panel structure of the data does not permit estimation of a model with a fixed firm effect, since union status does not change within the sample period. It is possible that the efficiency of the regression could be improved by allowing for a more complicated covariance structure (for example, across firms and overtime). However, the unbalanced nature of the design makes this approach computationally burdensome and it has not been pursued.

TABLE 4—UNIONIZATION AND FIRM PERFORMANCE—BASIC RESULTS[a]

	Profitability				Growth		Productivity			
	ROI	*ROI*[b]	*ROI*[b,c]	*ROS*[b,c]	*GS*	*GS*[b,c]	*LSL*	*LSL*[b]	*LSL*[b,c]	*LVL*[b,c]
UNION	−2.50	−4.05	−4.38	−1.60	−3.23	−0.67	−0.02	−0.02	−0.02	−0.03
	(0.86)	(0.87)	(0.84)	(0.38)	(0.98)	(0.83)	(0.01)	(0.01)	(0.01)	(0.01)
BIG3	–	−0.32	−0.12	−0.05	–	−0.02	–	−0.17[d]	−0.06[d]	−0.18[d]
		(0.02)	(0.03)	(0.01)		(0.02)		(0.02)	(0.02)	(0.04)
COMP	–	−0.30	0.03	0.01	–	−0.10	–	−0.26	−0.04	−0.12
		(0.03)	(0.04)	(0.02)		(0.03)		(0.03)	(0.03)	(0.05)
CPUR	–	−0.03	−0.02	−0.02	–	−0.03	–	0.01	0.01	0.05
		(0.03)	(0.03)	(0.01)		(0.03)		(0.03)	(0.03)	(0.05)
EXPER	–	0.15	0.13	0.07	–	−0.12	–	0.08	0.05	0.13
		(0.03)	(0.03)	(0.01)		(0.02)		(0.02)	(0.02)	(0.04)
RMG	–	0.05	0.06	−.02	–	0.78	–	0.05	0.06	0.11
		(0.02)	(0.02)	(0.01)		(0.02)		(0.01)	(0.01)	(0.02)
INDG	–	0.01	0.11	0.07	–	−0.04	–	−0.19	−0.12	−0.06
		(0.08)	(0.08)	(0.03)		(0.08)		(0.07)	(0.07)	(0.12)
INDC4	–	0.03	−0.03	0.01	–	−0.00[e]	–	0.02	−0.06	−0.13
		(0.02)	(0.02)	(0.01)		(0.02)		(0.01)	(0.01)	(0.03)
INDIMP	–	−0.10	−0.20	−0.11	–	0.10	–	0.03	−0.08	−0.35
		(0.06)	(0.06)	(0.03)		(0.06)		(0.05)	(0.05)	(0.09)
MS	–	–	0.84	0.41	–	–	–	–	0.74	1.21
			(0.06)	(0.03)					(0.05)	(0.09)
MS^2	–	–	−0.57[d]	−0.29[d]	–	–	–	–	−0.01	−0.01
			(0.08)	(0.03)					(0.0001)	(0.0001)
LKL	–	–	–	–	–	–	0.23	0.23	0.23	0.47
							(0.01)	(0.01)	(0.01)	(0.01)
LML	–	–	–	–	–	–	0.48	0.49	0.49	–
							(0.01)	(0.01)	(0.01)	
R^2	0.050	0.121	0.188	0.198	0.019	0.369	0.882	0.889	0.896	0.605
SEE	25.07	24.15	23.23	10.39	28.60	22.99	0.210	0.203	0.197	0.362
d.f.	4644	4627	4623	4623	4644	4622	4642	4625	4621	4622

Source: Estimated using data described in Table 2.

[a] In addition to the variables listed, each equation includes a set of 2-digit SIC dummies, a time trend, and a set of time-industry dummy interactions.

[b] This equation includes the 3-digit SIC variables: fraction of sales in Northeast, South, West; average tenure; mean grade attended; percent female; percent black, mean age.

[c] This equation includes variables measuring the fraction of sales to components inside the firm's parent corporation.

[d] Indicates coefficient has been multiplied by 10^2.

[e] Indicates less than 0.005.

difference in the average levels. On the productivity side, union firms have a slight edge (1 percent) in value-added per employee and a somewhat larger advantage in sales per employee; although firm conclusions await statistical analysis, it appears that the productivity differences may reflect differences in capital and materials per employee.

Estimates of various specifications of equations (21)–(23) are presented in Table 4, with a dummy variable (equal to 0 if percent organized is zero and 1 otherwise) used to measure unionization. Results with three sets of control variables are examined. The first (control 1) uses a set of 2-digit SIC dummies, a time trend, and a set of time-industry interaction terms.[36] The interaction setup allows each industry to have its own time trend, and is equivalent to using industry-level deflators. The second set (control 2) of

[36] Results with time dummies are very similar. Clearly, the cross-section variation in the data dominates variation over time.

controls adds all the variables on market structure and labor markets from Table 2, except the firm's own market share. The final set (control 3) adds market share and market share squared, and two variables which measure the fraction of sales or purchases made to or from components within the parent corporation (these variables control for differences in transfer prices).

In general, the evidence points to a sizeable negative effect of unionization on profits, and small or insignificant effects on other measures of performance. Consider first the union impact on *ROI*. The negative coefficient on the union dummy shown below suggests a decline of about 12 percent. Once additional controls for market structure and industry labor markets are added, however, the estimated effect increases to -4.1 percentage points, or a decline of about 19 percent relative to the sample mean. While the market share variables shown in the third column are highly significant, they have little effect on the union coefficient.

A negative union effect is also evident when profitability is measured by the return on sales. Only the results with the full model are presented (*ROS*, fourth col.), but the results with other specifications are similar. The size of the coefficient implies about an 18 percent lower return on sales in union firms. The size and significance of the *ROS* effect suggests that the decline in *ROI* arises primarily from a decline in total profits, rather than an increase in capital intensity. Indeed the similarity in the percentage changes in *ROI* and *ROS* implies that unionization has only a small effect on the ratio of sales to total capital.[37]

The large negative union effects in the profit equations are not repeated in the growth or productivity equations. Although significantly negative in *GS* (fifth col.) the union effect on the rate of growth relative to the market is small and statistically insignificant, once the characteristics of the market are introduced. The specification reported in *GS* (sixth col.) allows the market growth rate (*RMG*) to enter unconstrained. Estimates were also obtained with the *RMG* coefficient set equal to 1. The results were literally unchanged—the union coefficient was -0.67 with a standard error of 0.83.

The evidence in Growth implies that there is little difference in the way union and nonunion firms participate in the growth of the markets in which they operate. While this finding sheds some light on the effect of unionization, it does not rule out the possibility that unionization causes a one-shot decline in output relative to the market. In order to examine this possibility, real growth was replaced with the firm's market share as the dependent variable. The union coefficient in the specifications corresponding to *GS* (both cols.), were (with standard errors in parentheses): -0.016 (0.62), and 0.65 (0.52). When other measures of the firm's market structure were omitted, the union effect was -0.20 (0.64). There is thus no evidence of a union effect on the firm's share of its served market. Since market share reflects structural conditions in the market, the absence of a union effect may only suggest that union and nonunion firms operate in similar competitive environments. However, it would seem likely that a sizeable negative union effect on the firm's output would register in these data.

The analysis in Productivity of Table 4 turns to the question of the union productivity effect. Here, labor productivity is measured by sales and value-added per employee. Inclusion of *LKL* and *LML* in the equation means that the union coefficient measures differences in total factor productivity. As the results indicate, the choice of dependent variable has little effect on the union coefficient, although the sales specification in which materials and capital are included on the right-hand side is statistically superior. Unlike the results for profits and output growth, the union effect on productivity is unaffected by the inclusion of the control variables. Even though many of the structural variables are themselves significant, the union effect is estimated at -0.02

[37]This is in fact confirmed in the data. A regression with (S/K) as the dependent variable using the full model has a union coefficient of 0.09 (0.04). The mean of S/K is 2.29, which suggests a difference in S/K of about 4 percent.

(0.01) with or without them. When value-added is used, the union effect rises to -0.03 (0.01).[38]

While the union coefficient in the sales specification is twice the size of its standard error, it is substantively small; moreover, with over 4,600 observations, the power of the evidence that the effect is different from zero is not overwhelming. Yet these results stand in contrast to the results of Brown and Medoff, who found a positive union productivity effect using data on 2-digit SIC industries by state. Moreover, the results are contrary to some findings based on specific industries.[39] Reconciliation of these disparate findings is beyond the scope of the current paper, but it is possible to put the results here in some perspective. It must be remembered that the *PIMS* sample is not a random sample of firms in the manufacturing sector. Furthermore, the unit of observation here is not the establishment. The productivity measure refers to total employees of a business unit, and thus includes a variety of occupational categories not included (for example, sales, managers, engineers, accountants), or not included with the same weight in the establishment-based *Census of Manufacturers*. It should be noted, however, that when productivity equations were estimated controlling for the occupational mix in the firm's 3-digit SIC industry, no change in the results was obtained. It might be argued that failure to control for the blue-collar/white-collar mix would bias the union coefficient downwards, since unionized establishments are likely to be more blue-collar intensive and would therefore be somewhat less productive than white-collar intensive establishments (assuming white-collar workers are more productive). That argument, however, makes little sense within a 3-digit industry where the technology of production would seem to leave little scope for large differences in occupational mix.

Beyond the issue of comparabililty, it is important to note that the question of the union effect is ultimately an empirical question. There is no reason to suppose that the union effect has the same sign or magnitude in every industry. In fact, the average effect estimated in Productivity in Table 4 masks some differences across industries. Using the simple specification in the seventh column, estimates of the union effect were obtained for each 2-digit industry. While the majority of the industry effects (and the bulk of the sample) were quite close to the average of -0.02, positive union coefficients (6–17 percent) were found in textiles, furniture, and petroleum, and negative effects (4–18 percent) were found in lumber, stone, clay and glass, transportation equipment, and instrument manufacture. These estimates may reflect differences in the union-nonunion mix of component 3-digit industries, but they do suggest some diversity of effects. Evidence from previous industry case studies suggests that the nature of the union's impact on productivity depends on a complex interaction between management adjustment and union policy and action, which may well differ across industries.[40] On average, the interaction in the firms in this sample leads to little change in productivity, but the results suggest the need for further analysis of industry effects.

C. *Alternative Measures of Unionization*

Thus far, the analysis has been based on a relative simple specification of unionization. The results in Table 4 are in effect comparisons of conditional means for union and nonunion firms. It is possible to construct additional measures of unionization which may enrich understanding of its impact. Three possibilities are pursued in Table 5. The first is simply to enter the percentage of employees unionized in place of the union dummy. The second breaks up the percentage unionized variable into the three categories: less than 30 percent, between 30 and

[38]Some estimates were obtained with a union-time interaction term, but the overall results were little changed; the interaction term was not statistically or substantively significant.

[39]John Frantz (1976) and myself (1980) found positive union effects in the cement and furniture industry.

[40]The most explicit treatment of this issue is found in my 1980 article.

TABLE 5—ALTERNATIVE MEASURES OF UNIONIZATION[a]

Dependent	Profitability			Growth			Productivity		
Variable:	*ROI*	*ROI*	*ROI*	*GS*	*GS*	*GS*	*LSL*	*LSL*	*LSL*
%UN	−.06	–	–	0.002	–	–	−0.03^b	–	–
	(0.01)			(0.012)			(0.01)		
UN30	–	−4.23	–	–	−2.95	–	–	−0.02	–
		(1.54)			(1.52)			(0.01)	
UN60	–	−2.91	–	–	−0.62	–	–	0.01	–
		(1.06)			(1.04)			(0.01)	
UN60+	–	−5.28	–	–	−0.26	–	–	−0.04	–
		(0.93)			(0.92)			(0.01)	
Z30	–	–	−0.14	–	–	−0.08	–	–	−0.06^b
			(0.05)			(0.05)			(0.04)
Z60	–	–	−0.04	–	–	0.10	–	–	−0.01^b
			(0.06)			(0.06)			(0.05)
Z60+	–	–	0.04	–	–	−0.05	–	–	−0.03^b
			(0.05)			(0.05)			(0.04)
MS	0.83	0.84	0.83	–	–	–	0.01	0.01	0.01
	(0.06)	(0.06)	(0.06)				(0.0005)	(0.0005)	(0.001)
MS^2	−.005	−0.006	−0.006	–	–	–	−0.016	0.01	−0.01^b
	(0.001)	(0.001)	(0.001)				(0.001)	(0.001)	(0.001)
LKL	–	–	–	–	–	–	0.23	0.23	0.23
							(0.01)	(0.01)	(0.01)
LML	–	–	–	–	–	–	0.049	0.49	0.49
							(0.01)	(0.01)	(0.01)
R^2	0.187	0.189	0.188	0.369	0.369	0.370	0.896	0.897	0.896
SEE	23.24	23.22	23.23	22.98	22.98	22.98	0.192	0.196	0.197
d.f.	4623	4621	4621	4625	4623	4623	4621	4622	4619

Source: Estimated from data described in Table 2.

[a]Each equation includes the full set of control variables listed in Table 2, as well as those listed in fnn. a and b of Table 4.

[b]Coefficient has been multiplied by 10^2.

60 percent, and over 60 percent. These two measures are combined in the third approach through the use of a linear spline function. For the problem at hand, the spline variables are written as

$$
\begin{aligned}
Z30 &= U && 0 \leq U < 30 \\
&= 30 && U \geq 30 \\
Z60 &= 0 && 0 \leq U < 30 \\
&= U - 30 && 30 \leq U < 60 \\
&= 30 && U \geq 60 \\
Z60+ &= 0 && 0 \leq U < 60 \\
&= U - 60 && U \geq 60
\end{aligned}
$$

where U is the percentage unionized. The variables are constructed to eliminate discontinuities at the breakpoint, and to allow the union effect to vary with the percent organized, with different slopes in the different ranges. The coefficients on these variables measure the slope of the function over the specified range.

The alternative measures of unionization are examined in Table 5, using the full set of control variables (control 3) as defined in Table 4. It is evident that the basic results on profitability, growth, and productivity are little affected by the new forms of the union variable, although some interesting patterns emerge in the comparisons between measures. A large negative effect on profitability is found under all specifications, but the monotonicity of the effect implied by the use of percent organized is dubious. The spline function shows a strong negative effect from 0 to 30 percent, and then no significant slope

thereafter; the slope over the final range is actually positive. Use of three dummy variables suggests some differences depending on percent organized, but statistically, the differences are not particularly significant. A comparison of the standard errors of estimate for these equations and the comparable equation (*ROI*, third column) in Table 4 shows that the alternative measures of unionization provide little explanatory power beyond the union dummy.

Similar statements apply to the growth and productivity equations. Although some differences in the various coefficients appear, the data appear to have little or very weak information about any differential effect of changes in percent organized. As noted earlier, this is likely to be a reflection of measurement error in the basic unionization variable. In light of the evidence, the remainder of the paper focuses on results with a simple dummy variable to capture the effects of unionization.

D. *Implications of the Basic Results*

Seen in terms of the models developed in Section I, the results in Tables 4 and 5 appear to be inconsistent with the simple monopoly model with the firm on its demand curve. If the union wage effect in these data were in the proverbial 10–15 percent range, the profit results could imply an elasticity of substitution on the order of 2. Yet, with a union wage effect, growth and market share in the simple model should have declined substantially; instead, the evidence suggests little change. Since productivity also changes very little, there appears to be some support for the simple bargaining model in which only a wage effect is present.

Additional evidence on the bargaining interpretation of these results is provided by examination of the capital-labor ratio. Neither capital nor labor should deviate from nonunion values in the face of unionization if wages are bargained through a process of joint profit maximization as described in Section I. Thus, if the bargaining model is correct, the union coefficient in an equation explaining the capital-labor ratio should be zero.

Within the framework sketched out in Section I, the capital-labor ratio is determined by technology and relative prices, but is unaffected by market structure. In practice, it is important to control for differences in the market and industry, since the choice of technology is likely to be closely related to the types of products the firm manufactures. Evidence on the capital-labor ratio, therefore, is examined in the context of a model which includes controls for industry structure, technology, and labor market characteristics, but which excludes the firm's market share and the characteristics of its competitors.[41]

Table 6 presents estimates of the union effect on capital per employee under three specifications. The first includes only 2-digit industry dummies and industry-time interactions. The second adds controls for the labor market and the product at the 3- and 4-digit levels. The third specification adds a set of controls for the extent of vertical integration (ratio of purchases to sales and share of internal components in those purchases), in order to control for differences in technology in the chain of supply. Results were obtained with arithmetic and logarithmic versions of the dependent variable, and in both cases the union effect is not significantly different from zero, once the control variables are added. In line 1c, for example, the union dummy has a coefficient of -129.6 and a standard error of 89.1. The effect is also substantively small: compared to the sample mean of 3174, the union coefficient constitutes a decline of about 4 percent.

The difference in sign between the arithmetic and logarithmic versions suggested the possibility of outliers in the data. Examination of the residuals in equation 1b revealed a small number (a little over 1 percent of the sample) of observations with residuals far from the median. These were primarily chemical processing firms, a sector in which extremely large values of capital per employee would be expected. Estimation without the outliers had no material effect on the

[41] In effect, the industry variables play the role of 3- and 4-digit SIC industry dummies.

TABLE 6—THE EFFECT OF UNIONIZATION ON THE CAPITAL-LABOR RATIO

Dependent Variable	Union	Market Controls[a]	Vertical Integration Controls[b]	R^2	*SEE*	*d.f.*
1. *K/L* (arithmetic)						
a)	−515.1 (89.2)	–	–	0.252	2607	4644
b)	−168.1 (90.5)	X	–	0.311	2505	4630
c)	−129.6 (89.1)	X	X	0.338	2456	4627
2. *LKL* (logarithmic)						
a)	−0.048 (0.021)	–	–	0.267	0.614	4644
b)	0.030 (0.021)	X	–	0.315	0.595	4630
c)	0.036 (0.021)	X	X	0.352	0.579	4627

Source: Estimated from data described in Table 2.

[a] Includes all the variables in fnn. a, b, c of Table 4, except *BIG3*, *COMP*, and *CPUR*.

[b] Includes same variables as in fn. a, and ratio of purchases to sales, and fraction of purchases from components of the parent corporation.

magnitude of the coefficients. The conclusion that the union effect is not significantly different from zero is unaffected by discordant observations. Evidence on the capital-labor ratio thus provides further support for the bargaining interpretation of the union's impact on firm performance.

E. *Union Effects and Market Structure*

It is obvious from the evidence developed in Tables 4–6 that unionized firms earn substantially lower returns than their nonunion counterparts. As far as other measures of performance are concerned, however, unionization appears to have only a minor impact. While the evidence is thus not inconsistent with a bargaining model of unionization and firm performance, broader inferences about the union's impact on the allocation of resources must be viewed with some caution. It might be inferred, for example, that unions affect the division of total excess return between workers and the firm, but have little impact on the use of real resources. Yet even if that characterization were to apply to the sample average, there may well be some diversity in the pattern of effects across competitive situations which could have real effects over the longer run. It matters whether the average effect reflects a situation in which the union extracts substantial gains from firms with monopoly power, leaving the competitive sector little changed, or whether the bulk of the union impact falls on firms with little market power.

The analysis in Section I suggested a number of reasons why the union's impact might be different in different contexts, and, in fact, why the union may have a relatively larger impact on the smaller firms. Some insight into the effects of unionization in different environments is provided in Table 7.

Estimates of the basic model are presented for subsets of the overall sample defined on the basis of market share. High-market share firms are defined as those with more than 35 percent of the served market; the cutoff point for low-market share firms is 10 percent. This approach is relatively crude, since other aspects of the market may be important in determining the nature of competition and the firm's market power. A 10 percent share has a different meaning in a world where the

TABLE 7—UNION IMPACT IN DIFFERENT COMPETITIVE ENVIRONMENTS[a]

Independent Variables	Low-Market Share 1 (*ROI*)	2 (*GS*)	3 (*LSL*)	High-Market Share 4 (*ROI*)	5 (*GS*)	6 (*LSL*)
UNION	−4.69	0.96	−0.03	−0.34	−1.51	−0.03
	(1.39)	(1.66)	(0.01)	(2.10)	(1.90)	(0.02)
BIG3	−0.09	0.06	−0.01[b]	−0.45	−0.03	−0.12[b]
	(0.04)	(0.05)	(0.03)	(0.07)	(0.06)	(0.05)
COMP	0.01	0.06	0.01[b]	0.06	0.06	−0.12[b]
	(0.05)	(0.06)	(0.04)	(0.15)	(0.13)	(0.12)
CPUR	−0.20	−0.07	−0.12[b]	0.24	0.06	0.18[b]
	(0.05)	(0.07)	(0.05)	(0.08)	(0.08)	(0.07)
EXPER	0.18	−0.18	0.15[b]	0.24	−0.10	0.19[b]
	(0.04)	(0.05)	(0.04)	(0.07)	(0.06)	(0.05)
RMG	0.04	0.63	0.05[b]	0.11	0.82	0.10[b]
	(0.03)	(0.03)	(0.02)	(0.04)	(0.03)	(0.03)
INDG	0.17	−0.14	−0.32[b]	−0.24	−0.06	−0.14[b]
	(0.13)	(0.15)	(0.11)	(0.21)	(0.19)	(0.17)
INDC4	0.06	0.04	−0.08[b]	−0.08	−0.05	−0.13[b]
	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)
INDIMP	−0.33	0.41	−0.30[b]	−0.42	−0.10	−0.20[b]
	(0.13)	(0.16)	(0.12)	(0.15)	(0.14)	(0.12)
LKL	–	–	0.20	–	–	0.31
			(0.01)			(0.01)
LML	–	–	0.54	–	–	0.46
			(0.01)			(0.01)
R^2	0.120	0.263	0.913	0.220	0.455	0.895
SEE	21.39	25.43	0.191	24.74	22.32	0.198
d.f.	1385	1385	1383	894	894	892

[a]In addition to those listed, each equation includes a set of 2-digit SIC dummies; a time trend, a set of time 2-digit SIC interactions; and several 3-digit SIC variables: fraction of sales in Northeast, South, West; mean tenure; percent female; percent black; mean grade attended; mean age.
[b]The coefficient has been multiplied by 10^2.

principal competitors have 3 or 4 percent. While acknowledging the approximate nature of the approach, the use of market share should capture the basic tendencies in the data.

Estimates of the union effect on performance are presented in the first line of Table 7; Table 8 provides mean values of the performance measures. The results are revealing. Unionization has an extraordinary effect on the return on capital in low-market share firms. Taken at face value, the estimated coefficient (−4.7) implies something close to a 40 percent decline in profitability due to unionization. A similar decline emerges in estimates using the return on sales. This large decline in profits among low-market share firms stands in contrast to the results for firms with substantial market power. The estimated effect in column 4 is statistically and substantively insignificant. Whereas unionized firms with less than 10 percent share earn much lower rates of return, profits among the "rich and powerful" appear to be little affected by unionization.

TABLE 8—MEANS AND STANDARD DEVIATIONS (*SD*) FOR HIGH- AND LOW-MARKET SHARE FIRMS

Performance Measure	High-Market Share[a] ($N=940$) Mean	*S.D.*	Low-Market Share[b] ($N=1439$) Mean	*S.D.*
ROI	34.7	27.3	11.1	22.4
GS	11.8	29.5	6.8	29.1
LSL	3.93	0.60	3.80	0.63

Source: Calculated from data described in Table 2.
[a]Defined as more than 35 percent of the market.
[b]Defined as less than 10 percent of the market.

The contrast is not nearly as sharp for the other performance measures. There appears to be some difference in the growth equation,

with lower share firms having a positive union effect. But the coefficients are not significantly different. The estimated productivity effects are identical, although the coefficient for the high share firms has a larger standard error. For the low-market share firms, the pattern of effects across performance measures is much like the overall results in Table 4. Unionization is associated with a large impact on profits, and small or negligible effects on productivity and growth. Among high share firms, however, unionization has no statistically significant impact on firm performance.

Estimates were also obtained for the balance of the sample. The results revealed a lower union impact on productivity as market share increased, but the decline does not appear to be smooth. Among firms with market shares of 10–20 percent, the data suggest a union effect of −2.4 (1.7), or about 12 percent of the mean. The estimate for firms with 20–25 percent of the market was −5.8 (1.8), or 22 percent, quite close to the sample average. In view of the fact that the effect vanishes among firms with over 35 percent market share, it seems clear that the high-market share firms in Table 7 form a distinctive group.

The theoretical analysis in Section I provided several reasons why the impact of the union on profitability might be larger among firm's with less market power. Without information on relative wages (which is not available), it is not possible to draw firm conclusions, but the evidence appears to be consistent with studies which found little relationship between market structure and the union wage effect. In the context of the bargaining model, a constant union wage effect across market structure implies that unionization will fall more heavily on the less powerful firms.

While this line of thinking seems appropriate for most of the firms in this sample, a different explanation is likely to apply to those firms with substantial market power. The absence of a profit effect among high share firms could imply no union wage effect, but that is contrary to most of the evidence on the subject developed to date. A more appealing alternative relies on the possibility of limit pricing behavior. It does not seem farfetched to suggest that the firms in the high share group enjoy some barriers to entry (average market share = 50 percent) and may be limit pricing. If so, no profit effect would be observed even if unions raise wages, as long as the union effect was reflected in the limit price. If the limit price were below the monopoly price, and if the union imposes a cost which would be borne by new entrants, the limit pricing firm will adjust the limit price to cover the added costs, leaving the price cost margin unchanged.[42]

The evidence on market structure and union effects provides some perspective on the overall results, and suggests a number of interesting hypotheses about the union impact. The evidence on profitability is consistent with the models of competitive interaction developed in Section I. Unionization appears to bear most heavily on those firms whose profitability is already at a relatively low level. A disproportionate effect on less powerful firms raises the possibility that unionization influences processes of entry and exit and thus the structure of the market. This notion is not unreasonable. Oliver Williamson (1968) has argued precisely this point in the coal industry, and the effect of the UAW on the smaller automobile companies in the 1950's provides a further historical example.[43] This obviously depends on the extent of organization in the industry and the ability of new or smaller firms to avoid unionization.

It is difficult to pursue these hypotheses in the *PIMS* data, since the observations pertain to a single firm, while the hypothesis involves the extent of unionization in an industry. Because the data set is not necessarily composed of observations on direct

[42] See the paper by Martine Duchatelet and M. Therese Flaherty (1981) for a theoretical analysis of limit pricing which suggests that it may be a more pervasive phenomenon than previously believed.

[43] Robert Macdonald (1963) documents the effect on the so-called "Independents" of UAW wage policy. In this case, some of the smaller firms (i.e., Studebaker) actually paid wages above those at the larger firms. This "pattern-plus" approach to bargaining between the UAW and the smaller companies may have affected the "Independents" ability to compete and may have contributed to their eventual demise. See Macdonald, pp. 258–306.

competitors, and because little information about competitors is available (for example, the extent to which they are unionized), it is not possible to examine what are essentially industry dynamics. In this case, one is left with suggestive results about the incidence of the union impact on profits and a number of working hypotheses about longer-term effects on competition; the need for further research is clear.

III. Conclusions

This paper has developed a framework for analyzing the effect of unionization on firm performance, and has provided empirical evidence on several indicators, including rates of return on total capital and sales, output growth, market share, productivity and the capital-labor ratio. The analysis provides clear evidence that, on average, unionized firms earn substantially lower returns than nonunion firms operating in comparable technological and competitive environments. It is also evident that other dimensions of performance, particularly, growth and capital-labor substitution, are little affected by unionization in this data. The evidence is thus consistent with a bargaining model of union-firm interaction, in which the union affects the distribution of profits, but has little effect on output, or factors of production.

This characterization of the union's effect has some interesting implications. It explains why unionized firms may survive over long periods of time, and why the owners of capital would be strongly opposed to unionization. The evidence suggests further that the issue of the impact of the union on resource allocation is really a question about the long run. In this data, at least, efficiency effects would seem to derive from differences between firms with and without market power, and from differences between industries. Interpreted literally, the results point to factors which influence the maintenance of union organization, and exit and entry behavior as the key determinants of the union's impact on efficiency. One implication is that research on these problems could benefit from analysis of unionization in specific firms and industries with emphasis on the historical perspective. Such longitudinal analysis, if focused on a period of time long enough to allow for entry and exit, would be a valuable addition to the literature.

These comments are not meant to imply that the questions about the union's impact on the firm which motivated this study have been answered. The analysis is suggestive of a possible interpretation, but it is well to remember that the evidence in favor of the bargaining model is largely negative: the data fails to reject the notion that unionization has no effect on output or factor use. Furthermore, the special nature of the *PIMS* data suggests caution in embracing the bargaining model as generally applicable. Even within the data set there is some indication that the impact of the union may differ in different competitive settings. Further analysis of the impact of the union on the firm, of possible differences in productivity effects by industry, and of the role of the union in competitive dynamics within an industry seems warranted.

REFERENCES

Ashenfelter, Orley and Johnson, George, "Unionism, Relative Wages and Labor Quality in U.S. Manufacturing," *International Economic Review*, October 1972, *13*, 488–508.

Berstein, Irving, *Turbulent Years: A History of the American Worker 1933–1941*, Boston: Houghton-Mifflin, 1969.

Brown, Charles and Medoff, James L., "Trade Unions in the Production Process," *Journal of Political Economy*, June 1978, *86*, 355–78.

Clark, Kim B., "The Impact of Unionization on Productivity: A Case Study," *Industrial and Labor Relations Review*, July 1980, *33*, 451–69.

______, "Management Response to Unionization," Working Paper, Harvard Business School, 1983.

______ **and Griliches, Zvi,** "Productivity Growth and R and D at the Business Level: Results from the PIMS Data Base,"

Working Paper No. 916, National Bureau of Economic Research, June 1982.

Duchatelet, Martine and Flaherty, M. Therese, "Strategic Response to Entry Over the Product Life Cycle," unpublished paper, 1981.

Frantz, John, "The Impact of Trade Unions on Productivity in the Wood Household Furniture Industry," Honors Thesis, Harvard University, 1976.

Freeman, Richard B., "The Exit-Voice Trade Off in the Labor Market: Unionism, Job Tenure, Quits and Separations," *Quarterly Journal of Economics*, December 1980, *94*, 643–74.

______ **and Medoff, James L.,** "The Two Faces of Unionism," *Public Interest*, Fall 1979, *57*, 69–93.

______ **and** ______, "The Percent Organized Wage Relationship for Union and Non-Union Workers," *Review of Economics and Statistics*, November 1981, *63*, 561–72.

Galenson, Walter., *The CIO Challenge to the AFL: A History of the American Labor Movement, 1935–1941*, Cambridge: Harvard University Press, 1958.

Hall, Robert E. and Lilien, David M., "Efficient Wage Bargains under Uncertain Supply and Demand," *American Economic Review*, December 1979, *69*, 868–79.

Macdonald, Robert M., *Collective Bargaining in the Automobile Industry*, New Haven; London: Yale University Press, 1963.

McDonald, Ian M. and Solow, Robert M., "Wage Bargaining and Employment," *American Economic Review*, December 1981, *71*, 968–80.

Martin, Stephen., "Advertising, Concentration and Profitability: The Simultaneity Problem," *Bell Journal of Economics*, Autumn 1979, *10*, 639–47.

Phillips et al., Lynn W., "Product Quality, Cost Position and Business Performance: A Test of Some Key Hypotheses," Working Paper No. 83-13, Harvard Business School, 1983.

Pugel, Thomas A., "Profitability, Concentration and the Interindustry Variation in Wages," *Review of Economics and Statistics*, May 1980, *62*, 248–53.

Ravenscraft, David and Scherer, F. M., "The Lag Structure of Returns to R&D," unpublished paper, 1981.

Rosen, Sherwin, "Trade Union Power, Threat Effects and the Extent of Organization," *Review of Economic Studies*, April 1969, *36*, 185–96.

Segal, Martin, "The Relation Between Union Wage Impact and Market Structure," *Quarterly Journal of Economics*, February 1964, *78*, 96–114.

Seidman et al., Joel, *The Worker Views His Union*, Chicago: University of Chicago Press, 1958.

______, "Why Workers Join Unions," *Annals of the American Academy of Political and Social Science*, March 1951, *274*, 75–84.

Shephard, William G., "The Elements of Market Structure," *Review of Economics and Statistics*, February 1972, *54*, 25–36.

Slichter, Sumner, Healy, James and Livernash, Robert, *The Impact of Collective Bargaining on Management*, Washington: The Brookings Institute, 1960.

Warren-Boulton, Frederick R., *Vertical Control of Markets: Business and Labor Practices*, Cambridge: Ballinger, 1978.

Weiss, Leonard, "Concentration and Labor Earnings," *American Economic Review*, March 1966, *56*, 96–117.

Williamson, Oliver E., "Wage Rates as a Barrier to Entry: The Pennington Case in Perspective," *Quarterly Journal of Economics*, February 1968, *82*, 85–116.

Strategic Planning Institute, *PIMS Data Manual*, Cambridge: The Strategic Planning Institute, 1978.

U.S. Bureau of the Census, *1970 Census of Population: Industrial Statistics* (PC(2)-P7B), Washington: USGPO, 1973.

______, *1972 Census of Manufacturers*, Vol. II: *Industry Statistics*, Washington: USGPO, 1975.

A New Theory of Aggregate Supply

By ROGER E. A. FARMER*

Macroeconomic theory contains two competing explanations of business cycle fluctuations. According to the intertemporal substitution mechanism (*ITS*), observed variations in employment represent the optimal response of labor supply to (misperceived) variations in the expected real rate of interest. A competing theory is offered by the literature on overlapping contracts that attributes employment variations to contracted nominal wage rigidities. Such rigidities prevent prices from adjusting in a way that would allow quantities to mimic the theoretical predictions of the Walrasian market-clearing model. Both of these theories incorporate the property that employment converges in the long run to a natural rate of employment.

This paper develops an alternative theory. It presents a model of an economy in which agents are rational utility maximizers with perfect foresight of future prices—nevertheless the competitive equilibrium is Pareto inefficient. This situation arises because the economy does not contain a complete set of markets. Perfect capital markets do not exist because of asymmetric information (*AI*). A complete set of futures markets does not exist because agents have finite lives and trades cannot be negotiated with the unborn. These two sources of market failure are modeled by embedding a theory of *AI* contracts into an overlapping generations model (*OLG*) in which a critical role is played by the additional assumption that in some states of nature bankruptcy may be a binding constraint. The nonexistence of futures markets generates a model with the property that government debt is real wealth which implies that fiscal policy can have a permanent effect on the real rate of interest. The bankruptcy constraint translates changes in the rate of interest into changes in the frequency of contract default—simply stated, a high interest rate increases the probability that a firm will go bankrupt (or that it will layoff workers).[1]

In sharp contrast to either the new classical or the sticky price contract explanations of business cycle fluctuations, this theory is *not* a natural rate theory *even in the long run*. It predicts that the steady-state employment level is an endogenous variable which varies systematically with fiscal policy.

I. Model Structure

In a model with asymmetric information, it is important to be specific about who observes what when. It is convenient for this reason to use a modified form of the overlapping generations model which is adapted from the work of Peter Diamond (1965).

It is assumed that time is divided up into a series of periods and that during each period, agents *either* produce goods *or* they exchange goods. Periods of production are interspersed with periods of exchange and the two sorts of activities are referred to as pro-

*Assistant Professor of Economics, University of Pennsylvania, 3718 Locust Walk CR, Philadelphia, PA 19104. This paper is adapted from ch. 3 of my thesis "Macroeconomics and Equilibrium," submitted to the University of Western Ontario in January 1982. Support for later stages of the research is gratefully acknowledged from the SSHRC of Canada in the form of a research grant and from the Canada Council in the form of a Postdoctoral Fellowship. I thank Joel Fried, Michael Parkin, Peter Howitt, Nick Rowe, and participants at workshops at the universities of Western Ontario, Toronto, and McMaster for extensive comments on an earlier version of the work. The current version has benefited substantially from suggestions and comments from Costas Azariadis, David Cass, Karl Shell, Oliver Hart, Larry Epstein, Angus Deaton, and a number of anonymous referees of this *Review*, none of whom are to blame for any errors and omissions which may remain.

[1] In this paper, output variation is due to bankruptcy. My 1985 paper develops a model in which real interest rate fluctuations generate changes in the frequency of layoffs, but in both instances it is the existence of bankruptcy which causes contract inefficiency. See the discussion of this point in Section XI.

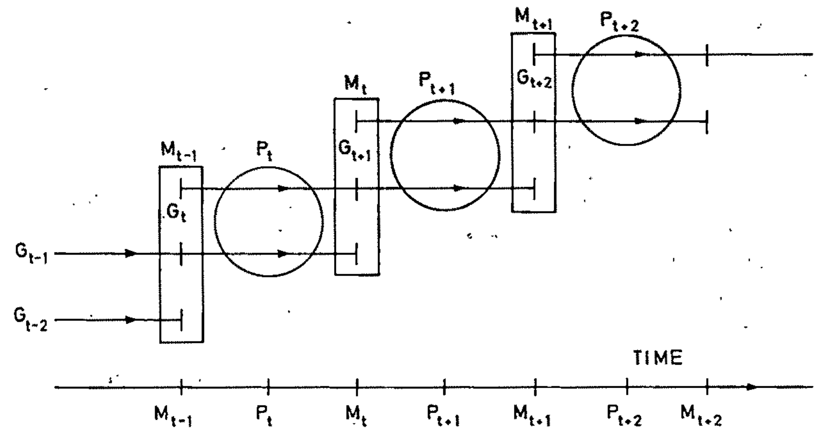


FIGURE 1

duction and market periods, respectively. Market periods are referenced with the notation M_t, $t=1,2\ldots$ and production periods with the notation P_t, $t=1,2\ldots$ and the ordered set $\{M_t, P_t\colon t=0,\ldots,\infty\}$ defines the sequence of all periods. Agents are assumed to live for two production periods and three market periods (see Figure 1 which outlines the temporal sequence of events).

I refer to generations with the notation G_t where the subscript t corresponds to the first production period of the representative agent's life. A representative agent of generation t is born in market period M_{t-1} at which there are three generations present—the newly born generation G_t, the generation G_{t-1} each of whom is in the transition from youth to old age, and the old generation G_{t-2}, each of whom is about to die. There is no growth and so each generation contains the same number of individuals.

In the first production period P_t, the individual supplies labor services that are combined with a single unit of capital. Labor and capital are combined according to the provisions of a contract that is written at the market period M_{t-1} immediately preceding production. It is assumed that production takes place on distinct "islands" and that there is no auction market for labor. There is a single good which may be consumed or used as capital in the following period. It is produced using a Leontieff technology and the gross quantity of output that is available on each island in each period (including depreciated capital) is given by

$$(1) \quad y_t = k_t(1-d-\Delta l_t) + \alpha \min(k_t, l_t)s.$$

The labor supply decision is dichotomous, $l_t \in \{0,1\}$. The worker chooses whether or not to work after observing a random productivity shock, s, and if he or she chooses to work then depreciation occurs at a faster rate than would otherwise be the case. The symbol d represents depreciation due to natural wastage, decay, etc., and Δ represents the additional depreciation that occurs if the machine is used for productive purposes.

In equilibrium, optimizing agents will invest a single unit of capital with each worker and (because of the assumption that the technology is Leontieff) this allocation will be independent of the rate of interest. The assumption of a fixed coefficient technology is restrictive, but nothing of substance hinges on it,[2] and it does result in a considerable clarification of the mechanism by which the rate of interest can affect the probability of bankruptcy. In Section V, I show that a higher real interest rate will cause a loss in output because more firms go bankrupt and because asymmetric information causes firms

[2]My 1985 paper presents a model in which the technology allows capital-labor substitution that has qualitatively similar results.

to write contracts in which bankruptcy may be inefficient. The assumption of a Leontieff technology allows us to abstract from the mechanism of capital-labor substitution (that would also suggest that a higher rate of interest would be associated with less output) and it serves to clarify the role of bankruptcy.

The random productivity shock s is assumed to be an independent drawing from a probability density function $h(s)$ with support in the interval [0,1]. Each worker receives a different drawing from the same density function. I refer to the wage as w_t and to the gross return to the capital owners (one plus the interest rate) as R_t.

At the market meeting M_t, the worker and the capitalist divide up the output y_t according to the provisions of the contract that was written at the previous market meeting. The worker consumes the amount c_t^0 and allocates the rest of his/her income across a portfolio of assets that may consist of contracts with the new generation of workers, or of holdings of government debt that pays a guaranteed real return $\overline{R}_t$ in the following period.

At the final market meeting of life, the individual consumes an amount c'_{t+1}, equal to his or her gross portfolio returns, and dies.

II. Risk and Time Preference

Nothing of substance in this analysis hinges on attitudes to risk and it is expositionally convenient for the presentation of a theory of contracts with default if both parties to the contract are risk neutral. The usual way of modeling preferences under uncertainty is to assume that agents maximize expected utility—but the assumption of lifecycle von Neumann-Morgenstern (N-M) utility functions links risk preferences and time preferences in an overly restrictive way.[3] To separate these factors I adopt the ordinal certainty equivalence (OCE) representation of preferences developed by Larry Selden (1978, 1979) as an alternative to the standard expected utility hypothesis.

At market meeting M_t, the agent will make choices over the certain-uncertain consumption pairs $c_t^0 \times c'_{t+1}$. At this point in time, wage uncertainty has been realized but the portfolio return is a random variable. The work by Selden allows one to define an ordinal utility index $u(c_t^0, \hat{c}'_{t+1})$ that expresses time preferences. $\hat{c}'_{t+1}$ is the certainty equivalent of the uncertain consumption stream $c_{t+1}(R_{t+1})$ that is found by defining a separate N-M utility functional to express the individual's attitudes to risk. We shall be concerned with the case in which agents are risk neutral. This implies that the agent will choose the bundle $c_t^0, \hat{c}'_{t+1}$ to maximize

$$u\left(c_t^0, E\left[\left(w - c_t^0\right) R_{t+1}\right]\right). \tag{2}$$

An implication of this specification is that the investor cares only about the mean of the distribution of the random portfolio return R_{t+1}.

The other signatory to a contract will be a newly born individual whose utility will indirectly depend on the wage w_t. In order to define how this agent will behave when faced with an uncertain prospect, I shall adopt a straightforward generalization of the OCE representation of preferences. It is assumed that the agent looks ahead to the choice that he or she will make in the following period and replaces c_{t+1}^0 with its certainty equivalent. This is found by defining a separate N-M utility functional to express attitudes to first-period consumption risk. It is also assumed that this functional is linear in c_{t+1}^0, which implies that the young agent will care only about the mean of the distribution of w_t.[4]

[3]The only N-M utility function that allows both parties to a contract to be risk neutral is of the form $\alpha + \beta c_t^0 + \gamma c'_{t+1}$; it implies that individuals have linear indifference curves.

[4]Formerly, it was assumed that at M_t the young agent maximizes $u(\hat{c}_{t+1}^0, \hat{c}'_{t+2})$ where

$$\hat{c}_{t+1}^0 = v^{0^{-1}}\left[\int_s v^0\left(c_{t+1}^0\right) h(s)\, ds\right];$$

$$\hat{c}'_{t+2} = v'^{-1}\left[\int_s v'\left(c'_{t+2}\right) h(s)\, ds\right].$$

Risk neutrality is imposed with the assumption that v^0 and v' are linear functions.

III. Aggregation Issues and Expectations

The introduction of noise at the micro level provides a motivation for the use of contracts, but it is inconvenient at the level of the economy as a whole. In order to sidestep some of the issues that are introduced by a stochastic rational expectations equilibrium, it is assumed that there exists a continuum of agents in each generation—agents are indexed by the random productivity shock that they receive in their youth. This assumption implies that aggregate output and relative prices will be nonrandom even though the output of each individual firm is a random variable. Aggregate output will be equal to the integral of the output of each individual firm, given by

$$Y=\int_0^1 y(s)\cdot h(s)\,ds. \tag{3}$$

In Section VIII, a rational expectations equilibrium will be defined. My assumption that there is no noise at the macro level implies that there is no aggregate uncertainty and so this equilibrium concept will reduce to the more tractable assumption of perfect foresight. Before describing the properties of a perfect foresight equilibrium, I describe in detail the sorts of contracts that would be written by rational agents in this economy.

IV. Contracts

Recent work in the theory of implicit contracts (Costas Azariadis, 1983; Sanford Grossman and Oliver Hart, 1981) has shown that asymmetric information induces employment distortions if the party with superior information is risk averse. Complete income insurance is precluded by the existence of moral hazard, but partial insurance is achieved at the cost of some productive inefficiency. The theory developed below is related to this work.

It is assumed that a contract is written between a risk-neutral worker/entrepreneur and a risk-neutral owner (or owners) of capital, and that it is the worker who has superior information. In place of risk aversion, assume that the worker has limited resources and so bankruptcy is a possible option in some states. This assumption causes the worker/entrepreneur to act as if he or she were risk averse and to write a contract in which employment is lower on average than it would be if both agents were symmetrically informed. An important implication of using a bankruptcy constraint to generate risk-averse behavior is that the magnitude of the employment distortion is predicted to vary systematically with the rate of interest.[5]

A contract is defined to be a payment to the owner of capital $R(s)$ and an employment level $l(s)$ as a function of the state of nature. Contracts that are written at market meeting M_{t-1} stipulate how the proceeds of production in production period P_t are to be divided between the worker and the capitalist. Assume that the worker must design a competitive contract in order to maximize his or her own expected wage, subject to the various constraints on the set of feasible contracts which are described below.

A. *Competition*

It is assumed that the worker/entrepreneur must bid for funds in a competitive contract market. Capital is rented from a member of the older generation who is assumed to have an alternative store of wealth which pays an expected return of $\overline{R}_t$. Since the capitalist is assumed to be risk neutral, a feasible contract must satisfy the constraint[6]

$$E[R_t(s)]\geq \overline{R}_t. \tag{4}$$

B. *Incentive Compatibility*

It may be shown that the allocation which results from an *AI* contract may always be duplicated by another contract in which it is in the interest of the agent to tell the truth

[5] Bankruptcy constraints are explored elsewhere in the literature by David Sappington (1983) and a related model at the micro level is developed by Joseph Stiglitz and Andrew Weiss (1981).

[6] Note that the assumption that the capitalist is risk neutral could be relaxed to allow risk-averse behavior. The assumption that there is no aggregate uncertainty would allow the individual investor to obtain complete income insurance through portfolio diversification and an individual risky contract would not offer a risk premium in equilibrium.

(Milton Harris and Robert Townsend, 1981; Robert Myerson, 1979). There will, therefore, be no loss to restricting attention to those contracts for which truth telling is an optimal strategy.

Since the worker may only choose two employment levels $l \in \{0,1\}$, it is clear that the capitalist will only be able to distinguish between two states. The set of feasible contracts will consist of those that make one payment R_1 if $l=1$, and another payment R_0 if $l=0$. *Ex post*, the worker will observe s and choose that employment level which maximizes his or her wage, where the wage will equal the value of the firm minus the payment to the capitalist:

$$(5)\quad w_t = (1-d)+\alpha s - \Delta - R_1 \qquad \text{if } l=1;$$

$$w_t = (1-d) - R_0 \qquad \text{if } l=0.$$

The entrepreneur will operate only if the first of these expressions exceeds the latter, that is, only if $\alpha s \geq \Delta + (R_1 - R_0)$.

C. *Limited Liability*

Since the worker has no initial wealth, payments to the capitalist can be no greater than the scrap value of capital. This places an additional constraint on the set of feasible contracts

$$(6)\qquad R_0 \leq 1-d.$$

This constraint turns out to be critical to the theory developed below.

V. The Optimal Contract

A Pareto optimal contract will solve the following problem:

$$(7)\quad \max_{R_0, R_1} (1-d) - R_0 + \int_{\alpha s = \Delta + R_1 - R_0} (\alpha s - \Delta - (R_1 - R_0)) h(s)\, ds,$$

subject to

$$(8)\quad R_0 + \int_{\alpha s = \Delta + R_1 - R_0} (R_1 - R_0) h(s)\, ds \geq \overline{R},$$

$$(9)\qquad R_0 \leq 1-d.$$

If the two parties to the contract were symmetrically informed, then the optimal employment rule would be given by

$$(10)\qquad l=1 \quad \text{if} \quad \alpha s \geq \Delta.$$

That is, they would produce only if the state is sufficient to cover the excess depreciation involved. If bankruptcy were not a problem then the *AI* contract could mimic this rule by choosing a fixed interest payment independently of the state. This contract would take the form

$$(11)\qquad R_1 = R_0 = \overline{R}$$

$$(12)\qquad l=1 \quad \text{if} \quad \alpha s \geq \Delta.$$

But if the worker has limited wealth and if this constraint is binding then the optimal contract will take the following form:[7]

$$(13)\quad R_0 = 1-d,$$

$$(14)\quad R_1 = \min\{R_1 | E[R(s)] = \overline{R}\};$$

where

$$(15)\quad E[R(s)] = (1-d) + (R_1 - (1-d)) \int_{\alpha s = \Delta + R_1 - (1-d)} h(s)\, ds,$$

$$(16)\quad l=1 \quad \text{if} \quad s \geq \Delta + R_1 - (1-d).$$

Here R_1 represents the gross return to risky debt that exceeds the expected market return in equilibrium because there are some states in which default occurs. Figure 2 is the graph of the expected value of a debt contract as a function of R_1.

Consider what happens to the value of a debt contract as R_1 is increased beyond the

[7]It is possible to show that this contract will also be optimal for a range of values of $\overline{R}$ if the worker is risk averse. In this case $R_1 > R_0$ even if bankruptcy is not binding.

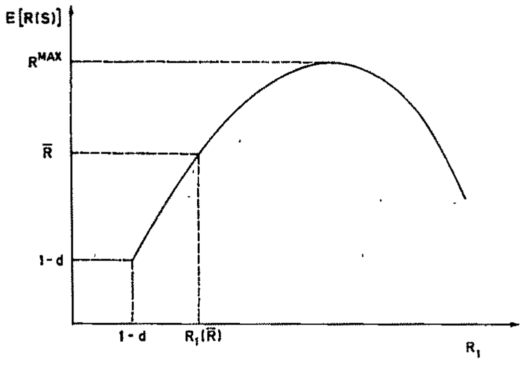


FIGURE 2

point $R_1 = 1 - d$ at which payment of the contract is guaranteed. Initially, raising R_1 will increase the expected value of the contract because the capitalist receives more in those states in which production occurs. But a higher value of R_1 is associated with a higher probability of bankruptcy and this effect tends to reduce the value of a contract. Eventually this second effect outweighs the former and the expected value of a contract attains a maximum which I denote R^{Max}. It is assumed sufficient regularity of the density function of s such that this maximum is unique.

This argument implies that there may be two values of R_1 for which the expected value of a risky contract is equal to $\overline{R}$ but the smallest of these values will Pareto dominate the other because it ensures a higher expected wage for the worker. For values of $\overline{R}$ in the range $[1-d, R^{\mathrm{Max}}]$, the gross interest payment on risky debt R_1 will be described as an increasing function of the return to alternative assets $\overline{R}$. This function is represented by equations (14) and (15) which I refer to with the shorthand notation

$$(17) \quad R_1 = R_1(\overline{R}) \quad \text{for} \quad \overline{R} \in [1-d, R^{\mathrm{Max}}].$$

VI. Aggregate Supply

If the rate of interest increases, then entrepreneurs will be forced to offer a higher return on contracts and the probability of bankruptcy in each location will be higher. But across the entire economy the full distribution of states will be realized and so an increase in the probability that each individual goes bankrupt will be translated into an increase in the observed frequency of contract failures. The aggregate supply of output in the economy is described by the following:

$$(18) \quad Y = (1-d) + \int_{\alpha s = \Delta + R_1(\overline{R}) - 1 + d} (\alpha s - \Delta) h(s)\, ds.$$

Notice that aggregate output decreases if R_1 increases. But R_1 is an increasing function of $\overline{R}$ that implies that for gross interest rates in the range $[1-d, R^{\mathrm{Max}}]$, aggregate output will be a decreasing function of the market rate of return.

$$(19) \quad Y = Y(\overline{R}) \quad \text{for} \quad \overline{R} \in [1-d, R^{\mathrm{Max}}],$$

where

$$(20) \quad dY/d\overline{R} < 0.$$

In a model in which the real interest rate varies systematically with government policy variables, this theory predicts that policy may have permanent effects on output and employment. The overlapping generations model is one framework which has these implications and in the remainder of this paper I explore the relationship between fiscal policy and aggregate supply that is generated by the mechanism described above.

VII. Asset Demand

Since the economy contains only one produced good, Walras' law implies that a description of asset market equilibrium is sufficient to characterize the equilibrium of the economy. At market M_t, each member of generation G_t solves the problem:

$$(21) \quad \underset{c_t^0}{\mathrm{Max}}\, u\left[c_t^0, \overline{R}_{t+1}(w_t - c_t^0)\right].$$

The solution will generate an asset demand function for each agent of the form

$$(22) \quad a_{t+1} = a_{t+1}(w_t, \overline{R}_{t+1}).$$

At M_t, the wage of each individual is no longer a random variable, since the produc-

tion shock has been realized. However, the distribution of wages across the population is nontrivial as each agent receives a different drawing from $h(s)$. Each agent's wage will be described as a function of the previous period's rate of interest and of the productivity shock in his or her specific location, found by subtracting the interest payment to the owners of capital from the value of the firm.

$$w_t = \max\{0, (\alpha s - \Delta - R_1(\overline{R}_t)\} \equiv \phi(s, \overline{R}_t). \tag{23}$$

The function ϕ is analogous to the factor price frontier described by Diamond in the context of an auction market economy, but the mechanism by which the interest rate affects the wage is different. To obtain an aggregate asset demand function, the individual functions must be integrated across the distribution of agents

$$A_{t+1} = \int_s a_{t+1}[\phi(s, \overline{R}_t), \overline{R}_{t+1}] h(s)\, ds \tag{24}$$

or

$$A_{t+1} = A(\overline{R}_t, \overline{R}_{t+1}). \tag{25}$$

Since R_1 is an increasing function of $\overline{R}_t$, each individual's wage will be a *nonincreasing* function of $\overline{R}_t$ that is strictly decreasing for all those individuals who receive a positive wage. (See equation (23).) Hence A_{t+1} will be decreasing in $\overline{R}_t$. It is also customary to assume that asset demands are increasing in the future interest rate $\overline{R}_{t+1}$, that is,

$$A_1 < 0; \quad A_2 > 0. \tag{26}$$

VIII. Asset Market Equilibrium

The supply of assets as a store of wealth is composed of the private stock of capital debt, K_t, and of the stock of government debt, B_t. The asset market will be in equilibrium when

$$A(\overline{R}_t, \overline{R}_{t+1}) = K_{t+1} + B_{t+1}. \tag{27}$$

But since there will be a single unit of capital inelastically supplied in each location (and a single unit in aggregate), this condition becomes

$$A(\overline{R}_t, \overline{R}_{t+1}) = 1 + B_{t+1}. \tag{28}$$

I define an equilibrium for the economy as a sequence of positive real numbers $\{\overline{R}^*_t\}_{t=0}^{\infty}$ and an associated sequence of asset demand functions such that (28) holds for all t. At market meeting M_t, the interest payment $\overline{R}^*_t$ is predetermined and so the equilibrium sequence $\{\overline{R}^*_t\}_{t=0}^{\infty}$ is completely determined by the sequence of government debt issues $\{B_t\}_{t=0}^{\infty}$. I focus on policy rules of the form $B_t = B$ for all t, and assume that the difference equation (28) converges to a unique stationary solution.[8] These assumptions allow me to unambiguously describe the effects of a change in policy regime on the stationary state of the economy.

IX. Policy Effectiveness

Figure 3 depicts an asset market equilibrium in the stationary state. The example is drawn for the case of a uniform density function and an asset demand function of the form $a = w \cdot \overline{R}/(1+\overline{R})$ but the qualitative features of this example are shared by a broad class of economics.[9]

The stationary asset demand function is increasing for low values of $\overline{R}^*$ as the positive effect of the future interest rate on asset

[8] This assumption is made by Diamond. The dynamics of the economy I describe are identical with the model described in his work and although we arrive at an aggregate asset demand function by a different route, its qualitative properties are similar to the nonstochastic economy which incorporates capital-labor substitution.

[9] This example was calculated numerically for values of $\alpha = 8$, $h(s) = 1$, $1 - d = \Delta = 0$. The individual asset demand functions are given by

$$a = R_{t+1}\phi(s, R_t)/(1 + R_{t+1})$$

where $\quad \phi = \max\{0, \alpha s - R_1(R_t)\}$;

$$R_1(R_t) = 1 - (1 - R_t/2)^{1/2}, \quad R_t \in [0,2].$$

The aggregate asset demand function is given by

$$A = R_{t+1}\left(2 - R_t/2 + 2(1 - R_t/2)^{1/2}\right)/(1 + R_{t+1}).$$

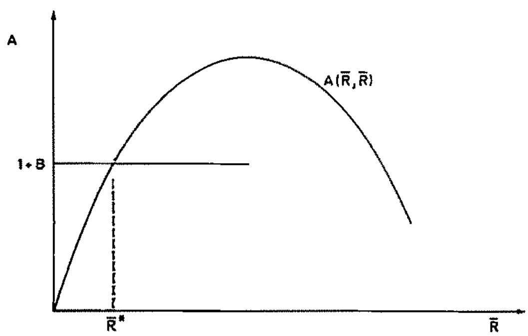


FIGURE 3

demand dominates the negative effect of the current interest rate on the wage. Eventually this function must slope down as asset holdings are bounded above by wealth that decreases with the interest rate. In order for an equilibrium to exist, the economy must be productive enough (the parameter α in the production function, equation (1), must be large enough) so that there exists an interest rate at which the old can be induced to hold the supply of assets. In terms of Figure 3, the stationary asset demand function must attain a value which exceeds $1+B$.

I define a fiscal policy in terms of the value of B. It is clear from this definition that changes in fiscal policy can affect the stationary equilibrium interest rate and it follows from the assumption that the economy converges to a stationary equilibrium that we can sign the direction of this effect. In the stationary state,[10]

$$dR^*/dB = 1/(A_1 + A_2) > 0. \tag{29}$$

Any policy that generates a higher real rate of interest will reduce output by increasing the frequency of contract failures, but, in an economy in which debt is net wealth, a policy that raises the real value of government debt will have exactly this effect. It follows that, in an economy in which debt is net wealth and some agents are better informed than others, fiscal policy will be able to permanently alter the steady-state frequency of contract failures.

X. The Government Budget Constraint

My description of policy in terms of the stock of debt has a counterpart in the more traditional definition of fiscal policy that refers to the market for flows. These two definitions are related by the government budget constraint

$$G_t = B_{t+1} - R_t B_t, \tag{30}$$

where G_t represents government purchases of goods.

In the stationary state this constraint takes the form

$$G = B(1-R). \tag{31}$$

This constraint is simpler than the government budget constraints that are encountered in most macroeconomic models because I have not explicitly modeled a monetary sector, nor included a source of tax revenues. Neither of these complications would significantly alter the conclusion that, in a model in which debt is net wealth, fiscal policy will have real effects.

I have deliberately avoided modeling taxes to avoid unnecessarily complicating the account of aggregate supply. (In a model in which bankruptcy is important, not even a lump sum tax will be nondistortionary because it will alter the probability of default.) As a result, the government cannot run a budget surplus in this model—G_t must always be nonnegative. However, it should be apparent that nothing of substance hinges on this issue and a model that allowed for taxes would have qualitatively similar properties.[11]

Another possible modification to the model would involve the addition of a motive for

[10] Around the stationary state, the time path of the interest rate will be approximated by the linear difference equation $dR_{t+1} = (A_1/A_2)\,dRt$, from which follows that $|A_2| > |A_1|$ if the economy converges. But since $A_2 > 0$, by assumption, this argument implies that $dR^*/dB > 0$.

[11] Note that if the interest rate is positive in equilibrium ($R > 1$), then the government must own some private capital, i.e., $B < 0$. This feature follows from the restrictive assumption of no taxes.

holding money that would permit a discussion of the relationship between inflation and aggregate supply. Money may be included in a variety of ways, but any method that has the property that only real balances matter will generate steady-state equilibria in which the rate of inflation equals the rate of monetary expansion. It follows that monetary hybrids of the model described in this paper are likely to have the property that any steady-state inflation rate is compatible with a variety of equilibrium levels of aggregate supply.

XI. Conclusions

The existing literature explains output variation over the business cycle in one of two ways. The equilibrium school attributes output fluctuations to voluntary variations in labor supply in response to perceived variations in intertemporal prices. The neo-Keynesian explanation hinges on prices that are slow to adjust. Both of these explanations are usually coupled with behavioral assumptions that generate the proposition that aggregate supply is independent of policy choices in the long run. In contrast, this paper suggests that output fluctuations are best described as movements of a rational expectations equilibrium in a model in which the equilibrium rate of employment is a function of rationally anticipated government policy rules. This suggests that excessive unemployment is a long-run phenomenon that will not go away if the economy is left to its own devices—if government policy has real effects, then it is important to choose the right policy. There are many possible criticisms of the arguments that have led to this position, some of which are discussed below.

It may be observed that the asymmetric information explanation of output fluctuations relies on the nonexistence of equity markets. This is true, but it is a feature explained within the context of the model by the more primitive assumption of asymmetric information. The contracts that are written under asymmetric information have many of the features of the debt contracts observed in the real world. I conjecture that partially relaxing the *AI* assumption will generate contracts that exhibit features of both debt and equity, but that the relationship between the rate of interest and the efficiency of the contract will be similar to the mechanism described in this paper.

In addition to the absence of equity, my explanation relies on the absence of futures markets. It is this assumption that allows the government to influence the terms of intertemporal trade. Once again, this feature is generated by a more primitive set of assumptions—agents have finite lives and that there are no consumption externalities. This latter assumption is potentially controversial since Robert Barro (1974) has shown that by permitting interdependent utilities, it is possible to describe a model in which a bequest motive will provide a substitute for futures markets. The substance of this issue involves the ability of policy to affect the real rate of interest and there would seem to be plausible theoretical justification for adopting a number of views—at this point, the issue must be judged on the empirical content of the research programs within which the alternative candidate theories are embedded.

It may be pointed out that the *AI* theory hinges on the importance of bankruptcy but that most of the output variation over the course of the business cycle is accounted for by variation in the incidence of layoffs. This criticism relies on a restrictive interpretation of the theory. Since I have modeled the entrepreneur and the workers as the same individual, I have been forced to model bankruptcy and unemployment as the same event. It is possible to construct a model (see my 1985 paper) in which these roles are separated and in which a firm writes a debt contract with a creditor and a labor contract with a workforce. In this context it may be demonstrated that an increase in the rate of interest causes the firm to write labor contracts that involve a higher probability of layoffs in every state of nature.

To my knowledge, the predictions of the *AI* model do not contradict any set of well-documented stylized facts although careful empirical work will be required in order to judge its consistency with the data. Recent work by Christopher Sims (1980) suggests that the interest rate is strongly causally prior to industrial production with a sign which is

consistent with *AI* theory, but this evidence must be regarded as tentative. Superficially, asymmetric information would appear to contradict the intertemporal substitution hypothesis (*ITS*) that suggests that there is a positive relationship between the interest rate and labor supply. But *ITS* refers to an expected *future* rate of interest in contrast to *AI* theory that relates current output to the *currently* realized real rate. This distinction means that empirical work which attempts to test the *ITS* theory is not directly relevant to *AI* theory.

There are other explanations of aggregate supply that suggest that the interest rate should cause output—for example, a neoclassical explanation which relies on capital labor substitution will have some of the same qualitative predictions. But there is evidence to suggest that workers are constrained in their labor supply decision (John Ham, 1983) which is not well explained within an *ex post* market-clearing framework. The *AI* theory can account for this evidence and yet still retain the concept of market clearing in an *ex ante* sense.

There seems to be no consensus position on the issue of whether fiscal policy can permanently influence the real rate of interest although Robert Shiller (1980) discusses a related issue concerning the effectiveness of Fed interventions on alternative definitions of the real rate. In view of the central role attributed to the real rate of interest by a number of alternative theories, this issue would seem to warrant careful empirical investigation.

It is too early to say if *AI* theory will prove to outperform its competitors as an explanation of business cycle fluctuations. But the evidence that has been brought to bear on existing theory suggests that neither neoclassical nor neo-Keynesian explanations are consistent with the behavior of postwar U.S. time-series data[12]—perhaps the time has come to give careful consideration to an alternative.

[12] Recent examples of studies which fail to support the *ITS* mechanism include work by Joseph Altonji (1982) and Orley Ashenfelter and David Card (1982). Ashenfelter and Card conclude that *neither ITS* theories *nor* sticky price Keynesian models are consistent with time-series evidence.

REFERENCES

Altonji, Joseph G., "The Intertemporal Substitution Model of Labour Market Fluctuations: An Empirical Analysis," *Review of Economic Studies*, Special Issue 1982, *49*, 738–824.

Ashenfelter, Orley C. and Card, David, "Time Series Representations of Economic Variables and Alternative Models of the Labour Market," *Review of Economic Studies*, Special Issue 1982, *49*, 761–81.

Azariadis, Costas, "Employment with Asymmetric Information," *Quarterly Journal of Economics*, Supplement 1983, *98*, 157–72.

Barro, Robert, "Are Government Bonds Net Wealth?," *Journal of Political Economy*, November/December 1974, *82*, 1095–118.

Diamond, Peter A., "National Debt in a Neoclassical Growth Model," *American Economic Review*, December 1965, *55*, 1127–56.

Farmer, Roger E. A., "Implicit Contracts with Asymmetric Information and Bankruptcy," *Review of Economic Studies*, 1985, forthcoming.

Grossman, Sanford J. and Hart, Oliver D., "Implicit Contracts, Moral Hazard, and Unemployment," *American Economic Review Proceedings*, May 1981, *71*, 301–07.

Ham, John C., "Testing for the Absence of Constraints in a Life-Cycle Model of Labour Supply," mimeo., University of Toronto, 1983.

Harris, Milton and Townsend, Robert, "Resource Allocation under Asymmetric Information," *Econometrica*, January 1981, *49*, 33–64.

Myerson, Robert, "Incentive Compatibility and the Bargaining Problem," *Econometrica*, January 1979, *47*, 61–74.

Sappington, David, "Limited Liability Contracts Between Principal and Agent," *Journal of Economic Theory*, February 1983, *29*, 1–21.

Selden, Larry, "A New Representation of Preferences over 'Certain and Uncertain' Consumption Pairs: The 'Ordinal Certainty Equivalent' Hypothesis," *Econometrica*, September 1978, *46*, 1045–60.

______, "An OCE Analysis of the Effect of Uncertainty on Saving Under Risk Preference Independence," *Review of Economic Studies*, January 1979, *46*, 73–82.

Shiller, Robert J., "Can the Fed Control Real Interest Rates?," in Stanley Fischer, ed., *Rational Expectations and Economic Policy*, NBER Other Conferences Series No. 16, Chicago: University of Chicago Press, 1980, 117–68.

Sims, Christopher A., "Comparison of Interwar and Postwar Cycles: Monetarism Reconsidered," *American Economic Review*, January 1980, *70*, 250–57.

Stiglitz, Joseph and Weiss, Andrew, "Credit Rationing in Markets with Imperfect Information," *American Economic Review*, June 1981, *71*, 393–411.

Deterrence by Market Sharing: A Strategic Incentive for Licensing

By NANCY T. GALLINI*

The potential of an incumbent firm to retain its dominant position in a market is a focal point of the modern theory of industrial organization. A recent literature examines this issue for markets in which challengers threaten an incumbent's position by engaging in research for a new technology. Richard Gilbert and David Newbery (1982) show that an incumbent firm with a first-mover advantage will preempt potential entrants by inventing a new technology slightly earlier than would its rivals. In a stochastic *R&D* model with simultaneous research decisions, Jennifer Reinganum (1983) demonstrates that the dominant position may change hands for sufficiently large inventions since the entrant has more to gain from innovation than does the incumbent. In both papers, the only preemptive strategy considered for the incumbent is the investment in additional research. An alternative strategy for an incumbent firm, which has not been considered, is the sharing of its current technology with a rival through a license contract.

This paper shows that an incumbent firm may license its production technology to reduce the incentive of a potential entrant to develop its own, possibly better, technology. If the licensing contract leaves the potential entrant with its expected return from further research, it will have no incentive to engage in further *R&D* activity. An incumbent firm, threatened by the risk of successful research by market challengers, can therefore secure a market share by selling the right to use its lower-cost technology. In contrast to previous models in which *R&D* activity deters entry into the product market, firms are encouraged into the product market as a way of deterring them from *R&D* activity.

The private and social incentives for licensing and its impact on research activity are analyzed by William Nordhaus (1969) and Pankaj Tandon (1982) for competitive product markets. Nordhaus shows that where a single cost-reducing invention is available to research, licensing at the profit-maximizing license fee is equivalent to producing as a monopolist; therefore its introduction does not alter the decision to research. When the pre-innovation market is characterized by an incumbent and potential entrants, Stephen Salant (1984) shows that licensing *ex post* to innovation can alter the research decision as anticipation of the returns from future licensing encourages research by the entrant. As a counterpoint to the Salant result, I find that licensing *ex ante* to research may discourage further research by the entrant. Indeed, this strategic incentive for licensing may be so strong as to persuade the incumbent as well as the entrant to terminate research activity. To isolate this incentive for licensing from the one discussed by Salant, a simple model is constructed in which licensing of future technologies is not desirable.

Anecdotal evidence suggests that the strategic incentive does lead innovators in some markets to license their technologies. Examples of this information sharing are prevalent in markets where technological change is rapid. In the 1940's, Standard Oil of New Jersey discovered a process for synthetic rubber. It traded patent rights on this process for a synthetic oil production process discovered by Farben, a German rubber company. In both the synthetic oil and synthetic rubber markets, rival companies were

*Department of Economics, Scarsborough College, University of Toronto, Toronto, Ontario, Canada, M5S 1A1. I thank Jon Cohen and Ron Saunders for discussions that inspired this paper. I am also grateful to Yehuda Kotowitz, Andy McLennan, Jennifer Reinganum, Ralph Winter, Myrna Wooders, two anonymous referees, and seminar participants at the University of Toronto for helpful comments on earlier drafts.

encouraged to take out licenses and led to believe that their requests for licenses would be granted. F. L. Vaughan writes, "The apparent purpose [of licensing] was to discourage independent research to find out [Farben's and Standard's] progress towards synthetic rubber" (1956, p. 146). Standard and Farben remained dominant in synthetic rubber and synthetic oil until 1942 when a government indictment charged them with agreements "... to use their combined present and future patents to prevent others from manufacturing and selling better and cheaper oil and chemical products" (Vaughan, p. 156).

The model for analyzing the strategic incentive to license is outlined in Section I, and in Section II the conditions are determined under which the current technology is licensed in equilibrium. In Section III the purely cooperative (or monopoly) level of *R&D* activity is compared to the level under the mixed cooperative and competitive arrangement of licensing and the level under pure rivalry. The effects on the incentive to license of changing the simplifying assumptions of small numbers of rivals and distinct technologies are examined in Section IV. The one-draw-at-a-time research plan is altered to ask when an incumbent will choose licensing of its current technology over preemptive research.

I. The Model

The role of licensing as a strategic instrument is illustrated in a simple model of *R&D* with competing technologies. The market for a single product is considered. Access to the market is restricted to only two firms, who also have sole capability for researching and developing processes for the production of the good. A large number of potential processes or technologies exist that are completely specified by a single dimension: the unit costs of production. The number of distinct production costs that can be achieved is finite and, for illustrative purposes, taken to be three.

The activity of *R&D* is represented as a draw, at a cost of D dollars, from a known discrete distribution of technologies over the average production costs C. The average costs that the technologies can take on are $C_1 < C_2 < C_3$, and the probability distribution is given by $p_i = Pr(C = C_i)$, $i = 1,2,3$. Since a large number of technologies are available at each cost level, the firms can be modeled with close approximation as drawing the technologies with replacement. The cost level of the technology chosen by a researching firm is immediately revealed to both firms, but technological information can only be acquired through a licensing contract. Property rights on the discovered technology are protected by a patent; without patent protection, licensing would never take place since the licensee would terminate the contract and continue to use the information.

The research game, assumed to take place in stages prior to the opening of the market, is effectively instantaneous. This assumption, adopted elsewhere in the *R&D* literature (for example, see Reinganum, 1982), simplifies the model considerably without affecting the basic results.

The analysis begins after a sequence of draws has already taken place, leaving the two firms with technologies of different costs. Specifically, the low-cost firm has drawn a technology with costs C_2; the high-cost firm has a technology with costs C_3. Finally, it is assumed that the inventions are drastic in that the low-cost firm (the incumbent) takes all and the high-cost firm (the entrant) is momentarily out of the market.[1]

At the beginning of the analysis, the incumbent is thus endowed with a monopoly position in the market for the good. The incumbent then decides whether to sell rights to its technology at some license fee. When

[1]This assumption requires that the average costs of the three technology categories be sufficiently far apart that the innovator of the lower-cost technology can set the unconstrained monopoly price. In this paper only drastic inventions are considered in order to isolate the strategic incentive for licensing. In Salant, licensing takes place on a technology discovered in the future. The invention is sufficiently small that if an entrant patents the technology, both the incumbent and innovator can coexist in the product market. The incentive to license in his paper can be attributed to the gain in profits from the replacement of duopoly by monopoly production. Hence, a second incentive for licensing, apart from the one discussed in this paper, arises.

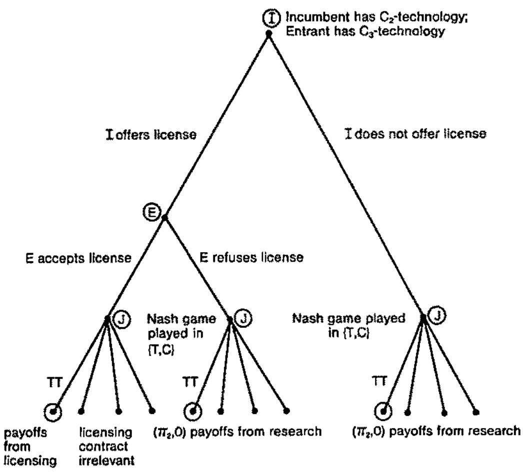


FIGURE 1. STRUCTURE OF THE INNOVATION GAME

Note: i = node where i decides, i = incumbent (I), entrant (E); j = "nodes" where both players choose a strategy from $\{T, C\}$ in Nash game, taking as payoffs the profits derived from future equilibrium strategies.

both firms face the same costs C_i, achieved either through a licensing contract or through research, they are assumed to share the cooperative profits.[2] These profits are denoted by Π_i. A licensing contract, therefore, leaves the entrant with a share of the industry profits equal to one-half the monopoly profits less some fixed license fee set by the incumbent. If an agreement is reached, research is terminated and production begins. If a license contract is refused or not offered, then the firms decide simultaneously between the actions continue (C) or terminate (T) research. The equilibrium concept we invoke is that of Perfect Nash Equilibrium (Reinhard Selten, 1975).

The game tree is illustrated in Figure 1. Although the firms move simultaneously in the subgame played in $\{C, T\}$, the incumbent in Figure 1 has been made the first mover in offering the license contract.[3] Hence, if there are gains from sharing, the incumbent will capture all of them by offering the rival its minimum acceptable share of profits. In fact, the entire interval between the entrant's minimum acceptable and the incumbent's maximum offered shares is determined in the model. I turn now to the conditions under which licensing is an equilibrium.

II. A Licensing Equilibrium

To isolate the strategic incentive for licensing current technologies on $R\&D$ activity, it is assumed that upon one firm's discovery of a C_1-technology, the rival firm would not have the incentive to draw another technology. This is described by the following condition:

$$(1) \qquad p_1\Pi_1/2 - D < 0.$$

This condition implies there is no incentive to license a C_1 technology to discourage further research by a rival. Furthermore, since the inventions are drastic, licensing will not yield rents in production from the replacement of a rival's relatively inefficient technology; with drastic inventions, the maximum profits are secured by producing solely from the low-cost technology.[4]

A. *Asymmetry in Returns from R&D*

Initially, an asymmetry is introduced in the returns from $R\&D$ for the entrant and the incumbent. In particular, it is assumed

[2]This assumption circumvents the issue of exclusive licensing discussed by Salant and is an assumption of convenience. The basic results would hold if the equilibrium in production were noncooperative (for example, Cournot) instead of collusive. The conditions for a licensing equilibrium in Section II would then describe relationships between the sum of the firms' expected returns from continuing research and the noncooperative industry profits under technology 2, rather than the cooperative profits.

[3]As an alternative, the $R\&D$ game could be modeled as sequential with the incumbent moving first. However, since there is already some built-in asymmetry in the pre-innovation market structure, it would then be difficult to identify whether the outcome should be attributed to the asymmetry in the market or in the $R\&D$ game.

[4]This would not be the case for small inventions. For small inventions, rents from the replacement of inefficient production technology could be recovered by licensing the low-cost technology to the high-cost firm.

that if the entrant engages in further research and discovers a technology with costs C_2, it will prefer to produce in this market and receive one-half the profits from production rather than continue to research alone until a lower-cost technology is discovered. This condition is described by

$$(2) \qquad \Pi_2/2 > \Pi_1 - D/p_1.$$

Condition (2) implies an asymmetry in research incentives due to two effects. First, the entrant, currently holding a high-cost technology, expects more from researching alone than does the incumbent since the entrant can benefit from discovering a technology with costs C_2. Second, under condition (2), the entrant has less incentive to stop researching when its rival continues than does the incumbent. If the entrant stops researching, it receives a zero payoff; if the incumbent terminates its *R&D* activity while the entrant continues to research, the incumbent faces the possibility of retaining one-half the monopoly profits from the current technology. Note that condition (2) is stronger than condition (1).

1. *The Alternative Equilibrium*. For sharing or licensing to be an equilibrium, two conditions must be met. First, each firm's profits under the license contract must exceed its profits from continuing alone. Otherwise, one firm would have the incentive to continue to research after the contract is signed. Upon discovery of the lower-cost technology, the contract would become irrelevant. Second, the licensing contract followed by no further research must leave the firms with profits at least as large as they would receive in the *alternative equilibrium*, the equilibrium in the Nash game played in $\{C,T\}$ if a licensing agreement were not reached. By not offering or refusing a contract, either firm can effect the alternative equilibrium.

To determine when licensing is an equilibrium in this research game, the payoffs must be defined under the alternative equilibria. Let R^I_{ij} and R^E_{ij} be the respective payoffs to the incumbent and entrant when the pair of actions $(i,j) \in \{C,T\}$ are actions taken at the no-licensing "node" of Figure 1, and given that the equilibrium strategies are followed beyond that stage. The payoffs from either or both firms continuing at the current cost pair (C_2,C_3) depend upon the expected equilibrium payoffs at different cost pairs. Since no further research takes place by either firm, upon discovery of a C_1 technology any cost pair that includes C_1 is mapped onto the pair of actions in which both firms terminate. If neither firm improves its technology in a draw, the firms keep their original technologies, and the equilibrium action at (C_2,C_3) is again adopted. Hence (C_2,C_2) is the only cost pair for which the equilibrium action remains to be determined.

The following argument shows that two potential pure strategy Nash equilibria at (C_2,C_2) must be considered: either both firms may continue to research (CC) or both firms may terminate research (TT). From (2), the best response by a firm to its rival's action of terminating research is also to stop and receive $\Pi_2/2$ rather than continue and receive the smaller expected profits of $\Pi_1 - D/p_1$. Thus, TT is always an equilibrium. However, if a firm continues to research at (C_2,C_2), its rival can either stop and receive 0 or compete for a C_1 technology and expect the returns

$$(3) \qquad R' = \Pi_1/2 - D/(p_1(2-p_1)).$$

If $R'>0$, both CC and TT will be Nash equilibria at (C_2,C_2); if $R'<0$, only TT is a Nash equilibrium—firms will terminate their *R&D* activity. TT is easily shown to be Pareto superior to CC and is therefore adopted as the outcome conditional upon (C_2,C_2).[5]

[5]I am grateful to an anonymous referee for pointing out the conditions for a CC equilibrium in the subgame starting at (C_2,C_2) in an earlier draft where the *R&D* game was sequential. In the sequential game for $R'>0$, the entrant would follow with the same action as the incumbent, and the incumbent, knowing this, would choose to terminate research since $R'<\Pi_1 - D/p_1$ and hence, $R'<\Pi_2/2$ by condition (2). In the simultaneous *R&D* game I restrict attention to pure strategies; because of the multiple equilibria a mixed strategy equilibrium may exist, but will not be persistent (Ehud Kalai and Dov Samet, 1982).

None of the results in this section depends on this selection.

Having determined the payoffs under all possible contingencies, the payoffs at the current cost pair (C_2, C_3) under the pairs of actions where both firms terminate research (TT), both firms continue to research (CC), and either the entrant or incumbent researches alone (TC or CT, respectively), can now be expressed as

$$(4)\quad TT:\quad R^I_{TT} = \Pi_2 \qquad R^E_{TT} = 0.$$

$$(5)\quad CC:\quad R^I_{CC} = R^E_{CC} = \frac{p_1(2-p_1)\Pi_1/2 + p_2(1-p_1)\Pi_2/2 - D}{1-(1-p_1)p_3}.$$

$$(6)\quad TC:\quad R^I_{TC} = \frac{p_2\Pi_2/2}{1-p_3}$$

$$R^E_{TC} = \frac{p_1\Pi_1 + p_2\Pi_2/2 - D}{1-p_3}.$$

$$(7)\quad CT:\quad R^I_{CT} = \Pi_1 - D/p_1 \qquad R^E_{CT} = 0.$$

For example, the denominator of the expected returns in (5) reflects the fact that if CC is the equilibrium pair of actions at the current cost pair, it will continue to be so as long as no discoveries are made, with probability $p_3(1-p_1)$. The numerator gives the expected return to either firm of one draw, given its rival also researches.

Only three of the four pairs of actions can, in fact, be alternative equilibria to licensing in the research game starting at (C_2, C_3): TT, CC, or TC. The incumbent researching alone is not a Nash equilibrium since the incumbent would prefer to terminate research rather than continue, given the rival also terminates. That is, a necessary condition for CT to be an equilibrium is $R^I_{CT} > R^I_{TT} = \Pi_2$, which contradicts (2). The conditions under which TT, CC, and TC are alternative equilibria are[6]

$$(8)\quad TT:\quad R^I_{CT} < R^I_{TT} = \Pi_2,$$

$$R^E_{TC} < R^E_{TT} = 0.$$

$$(9)\quad CC:\quad R^I_{CC} > R^I_{TC}, \quad R^E_{CC} > R^E_{CT} = 0.$$

$$(10)\quad TC:\quad R^I_{TC} > R^I_{CC}, \quad R^E_{TC} > R^E_{TT} = 0.$$

In sum, the payoffs of the no-licensing node (equations (4)–(7)) and the conditions characterizing the alternative equilibrium (equations (8)–(10)) have been derived. This sets the stage for determining when a licensing contract will be adopted by the rival firms.

2. *Existence of a Licensing Equilibrium.* As established in the previous section, a licensing equilibrium requires that a share of the profits of the current technology must exist such that both firms earn at least as much as they would (i) under the alternative equilibrium and (ii) from continuing research alone. These conditions for a licensing equilibrium will be met if the cooperative profits from the currently most efficient technology are at least as large as the sum of the firms' maximum profits from the alternative equilibrium or from researching alone. Let R^i_a be the payoff to firm i under the alternative equilibrium. Then, the two conditions for a licensing equilibrium can be summarized by the following condition:[7]

[6]The conditions in (8)–(10) are derived from the conditions for a Perfect Nash Equilibrium (PNE). For a PNE, the actions at any node of the game are equilibrium actions in the "one-stage" game with payoffs derived assuming equilibrium actions are followed at all future nodes. For example, if CC is an equilibrium pair of actions then the following must hold:

$$R^I_{CC} > p_2\Pi_2/2 + p_3 R^I_{CC} \quad \text{and} \quad R^E_{CC} > (1-p_1)R^E_{CC}.$$

That is, upon failing to draw a better technology, the equilibrium action of CC at costs (C_2, C_3) are followed at that node. The condition in (9) for a CC equilibrium, where the payoffs for each pair of actions are defined under the assumption that those actions are equilibrium actions (i.e., they are continually followed upon drawing technologies no better than (C_2, C_3)), are consistent with the PNE conditions. This can easily be shown for the other two equilibria.

[7]The condition for deterrence by sharing is that the monopoly profits from the licensed technology must exceed the sum of the expected duopoly profits from not

$$(11)\quad \Pi_2 > \max[R_a^I, R_{CT}^I] + \max[R_a^E, R_{TC}^E].$$

We need only consider whether a licensing contract is struck when market conditions are such that either CC or TC is an alternative equilibrium; when the alternative equilibrium is TT, current monopoly profits will not be shared since the entrant is not a threat.

Consider first the case when CC is the alternative equilibrium. In (11), R_a^I and R_a^E are replaced with R_{CC}^I and R_{CC}^E. But since CC is the equilibrium then $R_{CC}^i > 0$ and $\max[R_{CC}^I, R_{CT}^I] = R_{CT}^I$ and $\max[R_{CC}^E, R_{TC}^E] = R_{TC}^E$. Therefore, from (11), licensing is an equilibrium if

$$(12)\quad \Pi_2 > R_{CT}^I + R_{TC}^E.$$

Substituting the expressions for R_{TC}^E and R_{CT}^I from (6) and (7) into (12) gives

$$(13)\quad \Pi_2 > 2(\Pi_1 - D/p_1),$$

which is always fulfilled by condition (2). Thus, if CC is the alternative equilibrium, then a licensing arrangement is always reached. The range of shares to the entrant acceptable to both firms is $[R_{TC}^E, \Pi_2 - R_{CT}^I]$. Since the incumbent moves first, it will offer the minimum acceptable share R_{TC}^E.

Next, consider the conditions under which TC is the alternative equilibrium. Here R_a^I and R_a^E in (11) are replaced with R_{TC}^I and R_{TC}^E. The second max in (11) collapses to R_{TC}^E. But the first may equal R_{TC}^I or R_{CT}^I: from (6) and (7) note that R_{TC}^I may exceed R_{CT}^I for sufficiently low p_1 relative to p_2. When R_{TC}^I exceeds R_{CT}^I, (11) becomes

$$(14)\quad \Pi_2 > R_{TC}^I + R_{TC}^E.$$

Substituting the payoff expressions from (6) into (14) gives

$$(15)\quad \Pi_2 > \Pi_1 - D/p_1,$$

which always holds. The range of acceptable shares to the entrant are $[R_{TC}^E, \Pi_2 - R_{TC}^I]$. In the case where $R_{TC}^I < R_{CT}^I$, the condition for licensing is, as before, given by (13), which (again) always holds.[8]

In sum, *under asymmetry in incentives for research described by (2), reflecting the asymmetry in the production market, a licensing contract will always be struck to terminate research that would take place without licensing.* The incentives for the licensing contract can be analyzed either from the firms' collective point of view or from the incumbent's position alone (since the incumbent collects all the rents from the contract). From the incumbent's viewpoint, licensing protects against the risk of the discovery of a lower-cost technology by reducing the entrant's incentive for further research. From the firms' collective viewpoint, the expenditures that would be incurred without licensing would be excessive. The rents from the licensing contract reflect the savings of this excessive expenditure. The incentive for licensing may be so strong as to discourage the incumbent as well as the rival from further research; in the absence of licensing the incumbent would engage in research to lower the risk of being preempted by its rival for sufficiently high p_1 relative to p_2.

B. *Relaxation of the Asymmetric Returns Assumption*

Recall that condition (2) was imposed to reflect the possibility that the potential entrant has more to gain by investing in R&D than does the incumbent. As shown above, this condition always leads to a license con-

sharing. This is similar to the condition for preemptive patenting that is suggested in Gilbert and Newbery (G-N) where R&D investment is the strategic weapon. However, preemption occurs in our model and G-N for different reasons. In the G-N model, *deterrence by research* occurs, if the invention is not drastic, because the incumbent has more to lose by not researching than do the rivals. In this model, *deterrence by sharing* results from the condition that the incumbent has less to gain from continuing research than does the entrant.

[8] Recall that the Pareto-superior pair of actions TT was assigned to the cost pair (C_2, C_2). If $R' > 0$ (in (3)), then CC is a possible equilibrium. If CC were assigned to (C_2, C_2), the results in this section would be preserved. This is because the returns from continuing alone exceed those from both firms competing for a lower-cost technology.

tract that is Pareto superior to competition in $R\&D$. In this section, the firms are shown never to agree on a sharing arrangement if the condition in (2) does not hold. When condition (2) is false, the low-cost technology is relatively more attractive so that the entrant has the incentive to research until a C_1 technology is discovered. By not researching, the incumbent can expect to lose its entire market position. Since the gains from further research and from terminating research are identical for both the entrant and the incumbent, CT is a possible alternative equilibrium in addition to those given in (8)–(10). Whenever CT is an alternative equilibrium, TC will also be an equilibrium; however, the converse is not true.[9] The possible (pure strategy) alternative equilibria when (2) is false are TT, CC, TC, or both TC and CT.

Recall that for licensing to be an equilibrium the alternative equilibrium cannot be TT and the firms must receive the maximum of their payoffs in the alternative equilibrium or from researching alone. Since $R^I_{TC} = R^E_{CT} = 0$ when (2) is false and each firm would prefer to research alone than to compete with its rival, the payoffs from researching alone will never be smaller than the payoffs in the alternative equilibrium TC, CT, or CC. Hence, for licensing to be an equilibrium, a share for the entrant that satisfies (13) must exist. Condition (13), however, is identical to condition (2). This implies that if there is no asymmetry in the returns from research for the incumbent and entrant, there will not be any opportunities for effecting a Pareto-superior contract.

III. Implications of Licensing

In the stylized model of the previous section the incumbent and entrant always find licensing to be a profitable strategy when the returns from further $R\&D$ activity are greater for the entrant than the incumbent. Therefore, less research occurs in this mixed cooperative-competitive game than under pure rivalry. For symmetric returns, licensing never takes place in equilibrium.

In this section, the level of research activity undertaken by rivals is compared with that of a monopolist, facing the same conditions as the incumbent, but not threatened by entry. In particular, I ask whether rivals in an $R\&D$ market will terminate their research efforts and agree to license when an unthreatened monopolist would find further research unprofitable. The conventional wisdom is that competing researchers will invest at least as much as a monopolist.

The condition under which a protected monopolist will terminate $R\&D$ activity, having discovered a technology with costs C_2, is given by

$$\Pi_2 > \Pi_1 - D/p_1. \tag{16}$$

The monopolist's $R\&D$ activity given by this condition is compared with its strategy when suddenly faced with a potential entrant, first under the asymmetric returns condition (2), then under its complement that the expected returns from $R\&D$ are identical for the rival and incumbent.

Recall that when condition (2) is imposed, licensing always takes place. Furthermore, (16) is always fulfilled. Therefore, an unthreatened monopolist will terminate research in the asymmetric case as will rivals in the $R\&D$ market with opportunities for licensing.

When (2) is false, the lowest-cost technology is relatively profitable, so that current profits are not sufficient to impede further research. However, (16) may hold even when (2) is false and if so, the unthreatened monopolist will terminate research. Hence, under symmetric returns from $R\&D$, licensing cannot eliminate the investment in $R\&D$ when this investment is in excess of that undertaken by a monopolist.[10]

[9] That is, when CT is an equilibrium set of actions, then $\Pi_2 < \Pi_1 - D/p_1$. This implies that the expected payoffs from the entrant researching alone are positive ($\Pi_1 - D/p_1 > 0$) and hence, TC is also an equilibrium. However, when TC is an equilibrium, then $\Pi_1 - D/p_1 > 0$ but this is consistent with the condition for the incumbent to stop researching when its rival terminates: $\Pi_2 > \Pi_1 - D/p_1$.

[10] One may be bothered by the presumption that firms can agree on their production but cannot coordi-

The above analysis implies that licensing may reduce the extent of *R&D* in excess of that undertaken by a monopolist (or a social planner facing the same rewards from invention): when the returns from research are asymmetric, licensing is adopted whenever a monopolist finds that the benefits from pursuing better technologies are less than the research costs. There is yet another waste from rivalrous activity in an *R&D* market that may be reduced by licensing. This is the waste from duplicated efforts that arise when both firms find research activity profitable. For example, suppose both firms start the *R&D* game with C_3 technologies. Furthermore, let the remaining technologies be sufficiently profitable that both firms would simultaneously draw technologies until a superior one was discovered in the absence of licensing. When the research periods are short, this simultaneous research is wasteful since it may result in redundant discoveries and, hence, does not minimize the expected costs of an invention. In this case, a single researcher, finding further research profitable, would minimize research costs by drawing one technology per period.

If licensing were to be an equilibrium upon the discovery of a C_2 technology, then one firm may have the incentive to wait, letting the other firm research. The passive firm risks the chance of its rival drawing a C_1 technology and some of the returns from the C_2 technology, if one is discovered, because of the license fee it must pay. But if the probability of discovering a sharing technology is large compared to that for the lowest-cost technology and if the equilibrium license fee is sufficiently small, the Nash equilibrium under licensing may be characterized by one researching firm.[11]

In summary, licensing may reduce investment in *R&D* to levels close to those set by a protected monopolist in two ways. First, when an unthreatened monopolist terminates research upon discovery of a technology of a particular quality (defined by the unit costs of production), that same technology may be licensed to a rival threatening to engage in *R&D*. Second, licensing may encourage one of the firms to wait if its rival researches, given the expectations that a sharing offer may be imminent. Inefficiencies from parallel research efforts (as opposed to sequential research efforts) may be avoided.

nate *R&D* activity. However, evidence suggests that in small numbers situations, firms have more difficulty coordinating nonprice instruments than agreeing on price (F. M. Scherer, 1970).

[11] The availability of licensing has an opposing effect from the one discussed above on the incentives to research in the early stages: licensing may *encourage* simultaneous research since it increases the payoffs from research. Since the first firm to discover a "licensing technology" can persuade its rival to terminate research, it will receive larger profits from this outcome than when licensing is not an option. The net effect on the incentives for both firms to engage initially in research can be analyzed under the condition in (2). In this case, a firm that discovers a C_2 technology, while its rival stays at the status quo, will offer an acceptable license contract, specifying Π_2^L for its rival and $\Pi_2 - \Pi_2^L$ for itself. When licensing is not an option, the high-cost firm will continue to research until it also discovers a C_2 (or better) technology, leaving the first innovator with expected payoffs $((p_2\Pi_2/2)/(1-p_3))$. The payoffs from the other research outcomes are the same for the licensing and no-licensing cases. Hence, the expected payoffs to each firm from simultaneous research under licensing, R_s^L, and in the absence of licensing, R_s, are related by

$$\text{(a)}\quad R_s^L = R_s + \delta\left[\left(\frac{\Pi_2}{2} - \Pi_2^L\right)(1-p_3) + p_1\frac{\Pi_2}{2}\right],$$

$\delta = p_2 p_3/(1-p_3)^2(1+p_3)$. Since $\Pi_2^L \leq \Pi_2/2$, $R_s^L > R_s$. First consider the case discussed above where $R_s^L > R_s > 0$. In the absence of licensing, both firms will compete for a lower-cost technology in the early stages. Under licensing, both firms may not research since an alternative of waiting for a possible license contract is available to one of the firms. The expected payoffs from waiting are

$$\text{(b)}\qquad R_w = p_2\Pi_2^L/(1-p_3).$$

From (a) and (b) it is evident that licensing will encourage one firm to remain passive in the early research stages when the payoffs to the potential licensee are large and the probability of discovering a C_2 technology by the researching firm is high. Next consider the case where $R_s < 0$. When licensing is not available, only one firm engages in *R&D* activity in the initial stages. Note, however, that R_s^L may be positive for large license fees (small payoffs to the licensee) and a high probability of drawing a licensing technology when both firms research $(p_2 p_3)$. Under these conditions, R_s^L may exceed R_w. Hence, the higher returns from research due to the availability of licensing may actually encourage the firms to engage in parallel research efforts in the early stages.

IV. Comments on the Simple Model

While the simple model above is sufficient to illustrate the strategic incentive for licensing, some restrictions of the model may raise questions as to the general effectiveness of this alternative to preemptive research. The implications of relaxing three assumptions of the model—the small number of firms, the small set of distinct technologies, and the one-draw-at-a-time research plan—are briefly discussed in this section. Relaxation of the last assumption provides a natural framework for contrasting preemptive patenting with strategic licensing.

A. *Increasing the Number of Rivals*

The small number of rivals is the most restrictive feature of the model. When $p_1\Pi_1 - D > 0$, licensing can never take place in this model with free entry. If an entrant is persuaded to license a C_2 technology in this case, then the gains from entering the *R&D* market to discover a C_1 technology still exist for another entrant. Hence, as long as firms license the technology, there is a threat of entry and no licensing will take place.

When $p_1\Pi_1 - D < 0$, however, a licensing contract may be struck with free entry. For example, a single entrant into *R&D* activity may discover either a C_1 technology with probability p_1 and capture the entire monopoly profits or a C_2 technology with probability p_2 and (under condition (2)) share the market with the incumbent. This yields expected payoffs from a draw of $p_1\Pi_1 + p_2\Pi_2/2 - D$. Suppose these payoffs are positive and a license contract is struck. A second entrant may not have the incentive to enter *R&D* activity since the returns from discovering a C_2 technology would be reduced with three producing firms.

A simple example illustrates this possibility. Let $C_1 = 0$, $C_2 = .5$, $p_1 = .2$, $p_2 = .8$, and $D = .07$. The demand curve is given by $Q = 1 - P$ and derived monopoly profits from technologies 1 and 2 are $\Pi_1 = .25$ and $\Pi_2 = .0625$. The incumbent's expected payoffs from a draw, given it has a C_2 technology, are $p_1\Pi_1 - D = -.02$. A single entrant, facing the additional possibility of discovering a C_2 technology and sharing Π_2, expects the payoffs of .005 from a draw. A second firm will not have the incentive to research since it expects $-.01167$ from researching simultaneously with its rival. Suppose the incumbent and the entrant agree to a licensing arrangement and the entrant cancels further research plans. A second potential entrant may still find that research is unprofitable even if it is the only firm engaged in research. If it receives one-third of the market upon discovery of a C_2 technology, then its expected payoff from one draw is $-.0033$. Hence, a license contract is struck with the first entrant and further entry is unprofitable.[12]

B. *Increasing the Number of Cost Categories*

If the potential technologies remaining at the time of a research decision exhibit a wide range of distinct production costs that are lower than the status quo, a licensing contract may be difficult to agree upon. Further research may take place, eventually resulting in licensing of a lower cost technology or the discovery of a reservation technology, beyond which no further research is profitable.

When several types of technologies are introduced and the probabilities of these are high, the incentive for both firms to engage in research activity increases. However, if a

[12] For different parameter values, a second potential entrant may find research profitable when its rivals enter into a licensing agreement. For example, let $D = .065$ and the remaining parameterization in the previous example be the same. As before, only one firm initially threatens to research, expecting payoffs of .01 per draw. However, if a license contract is struck between the incumbent and this firm, a second entrant will threaten to research, expecting .0017 from a research draw. Knowing this, the incumbent and first entrant will agree to a license contract only if they can be guaranteed their expected payoffs from the alternative equilibrium in which only the first entrant researches. Since the monopoly profits from the current technology (.0625) exceed the sum of the expected profits to the incumbent and the first entrant from the alternative equilibrium (.025 + .01 = .035) plus the expected payoffs to the second entrant from researching (.0017), the three firms will enter into a licensing agreement. In cases where the profits from the current technology are not sufficiently large to accommodate the second entrant, no licensing agreement is reached and the first entrant researches in equilibrium.

large set of those technologies are ones for which licensing or sharing will subsequently be an optimal strategy, then one firm may want to wait for the outcome of its rival's endeavors. This is one of the savings that may be realized from licensing, as discussed in the previous section.[13]

C. *Licensing as an Alternative to Preemptive Research*

This paper is motivated, in part, by the possibility that an incumbent may use licensing rather than preemptive research to maintain its dominant position (or to avoid losing its entire market share). However, this conjecture cannot be analyzed fully in the simple model because firms draw only one technology per period. While this pattern of research does indeed represent the cost-minimizing approach when a firm is uncontested in $R\&D$ prior to market production, firms may invest in more than one draw when faced with rivals.[14] To compare preemptive researching with licensing, consider the following changes in the simple model. First, let more than one technology be drawn per instant, and second, let the incumbent be the first-mover in its research strategy. (The latter change is consistent with the incumbent being the first-mover in offering the license contract.) Under these assumptions, there will exist some number of projects, N^* for the incumbent, such that the entrant's expected return from research is zero.[15] If the incumbent draws N^* technologies per instant, it will completely deter entry and its expected profits will be $\Pi_I^* = ((1-(1-p_1)^{N^*})\Pi_1 - N^*D)/(1-(1-p_1)^{N^*})$. Alternatively, if $\Pi_2/2 > \Pi_1 - D/p_1$ (condition (2)), the incumbent may prefer to license its technology and receive at least $\Pi_2/2$. A necessary condition for preemptive researching to be preferred to sharing the current market in this case is for $\Pi_I^* > \Pi_1 - D/p_1$. This condition never holds for $N^* \geq 1$, suggesting that under (2) licensing is a still more profitable alternative to complete preemptive researching when many research avenues can be investigated simultaneously.[16]

V. Conclusions

This paper identifies circumstances in which firms engaged in $R\&D$ have the incentive to reveal information about their discoveries. In a small numbers situation, it is asked under what conditions will firms license their technologies to potential researchers. One answer to this question is provided by a model of process inventions in which a number of technologies with several different costs of production are available for discovery through research. Incorporation of this feature leads to a prediction of strategic behavior on the part of a firm currently dominating the market. The incumbent firm must make a decision to offer its technology to the rival, or risk being pushed out of the market. Thus, we have a theory of strategic deter-

[13] The effect of licensing current and future technologies on research activity is analyzed in a model with a large number of cost categories in a paper by myself and Ralph Winter (1984). The two incentives for licensing discussed in fn. 1 are combined in the paper (with some modifications) to analyze the overall impact of licensing on $R\&D$ activity.

[14] For a license contract to be acceptable, recall that each firm has to receive at least the (maximum) expected profits from researching alone. Because the research process is timeless, these profits are based on a one-draw-at-a-time research plan.

[15] Alternatively, the incumbent may try to partially deter the entrant; i.e., it may attempt to decrease the amount of research undertaken by the entrant but not necessarily to zero. Only complete preemption is considered here.

[16] For this comparison to be valid, it must be the case that the results on licensing in Section II are preserved when the incumbent is allowed to move first in its choice of research. In the case where the incumbent is the first-mover in both research and in offering the license contract, the possible alternative equilibrium under (2) are the same as before (CC, TC, TT) and CT. When CT is the alternative equilibrium, we have the same situation as described by (12) and (13) in Section II. When CC is the alternative equilibrium, the incumbent may engage in preemptive research; however, the returns from doing so will be smaller than R_{CT}^I. Hence, licensing always occurs for asymmetric returns in the sequential game. When condition (2) does not hold, a licensing contract is never struck but preemptive research may be undertaken.

rence by inducement of entry.[17]

A testable hypothesis that evolves from the *R&D* literature is that if patent protection is guaranteed to an innovator, more research will take place in a market with rivalry than in monopolistic *R&D* markets. This paper supports this prediction only when the returns from research to the firms are symmetric. Under asymmetry in returns to research, reflecting an asymmetry in market positions, this model predicts that the successful researcher attempts to secure some of its market position by licensing its current technology to its rivals. Research investment, beyond that undertaken by an unthreatened monopolist, is not undertaken by rivals when licensing is available. Indeed, a license contract may lead to termination of research even when the undiscovered technologies are sufficiently profitable that both firms could be accommodated in *R&D* activity. Furthermore, this model predicts that "wasteful" *R&D* from simultaneous research by rivals may be lessened under licensing, since a firm has an increased incentive to wait for a possible offer to share a technology discovered by its rival.

[17]This theory can explain a range of economic phenomena wider than those in industrial research and development. Large, more efficient firms are often observed to tolerate or even encourage less efficient rivals already in the market. An explanation commonly offered for this behavior is compliance with the antitrust laws. An alternative explanation, suggested by this model, is that a dominant firm would prefer to keep a familiar rival in the market than to face the uncertain outcome of the rival's efforts to survive in the market or from replacement of the rival by another, possibly more efficient, firm.

REFERENCES

Dasgupta, Partha and Stiglitz, Joseph, "Uncertainty, Industrial Structure, and the Speed of *R&D*," *Bell Journal of Economics*, Spring 1980, *11*, 1-28.

Gallini, Nancy and Winter, Ralph, "Licensing in the Theory of Innovation," Working Paper, University of Toronto, 1984.

Gilbert, Richard and Newbery, David, "Preemptive Patenting and the Persistence of Monopoly," *American Economic Review*, June 1982, *72*, 514–26.

Kalai, Ehud and Samet, Dov, "Persistent Equilibria in Strategic Games," Discussion Paper No. 515, Northwestern University, 1982.

Lee, Tom and Wilde, Louis L., "Market Structure and Innovation: A Reformulation," *Quarterly Journal of Economics*, March 1980, *94*, 429–36.

Loury, Glen C., "Market Structure and Innovation," *Quarterly Journal of Economics*, August 1979, *93*, 395–410.

Nordhaus, William D., *Invention, Growth and Welfare*, Cambridge: MIT Press, 1969.

Priest, G. L., "Cartels and Patent License Arrangements," *Journal of Law and Economics*, October 1977, *20*, 309–78.

Reinganum, Jennifer, "Strategic Search Theory," *International Economic Review*, February 1982, *23*, 1–17.

______, "Uncertain Innovation and the Persistence of Monopoly," *American Economic Review*, September 1983, *73*, 741–48.

Salant, Stephen, "Preemptive Patenting and the Persistence of Monopoly: Comment," *American Economic Review*, March 1984, *74*, 247–50.

Salop, Steven, "Strategic Entry Deterrence," *American Economic Review Proceedings*, May 1979, *69*, 335–38.

Scherer, F. M., *Industrial Market Structure and Economic Performance*, Chicago: Rand McNally, 1970.

Selten, Reinhard, "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, January 1975, *4*, 25–55.

Tandon, Pankaj, "Optimal Patents with Compulsory Licensing," *Journal of Political Economy*, June 1982, *90*, 470–86.

Vaughan, F. L., *The United States Patent System*, Norman: University of Oklahoma Press, 1956.

Adaptive Responses to Chemical Labeling: Are Workers Bayesian Decision Makers?

By W. KIP VISCUSI AND CHARLES J. O'CONNOR*

A fundamental issue in the economics of uncertainty is how individuals process information and make choices under uncertainty.[1] In a recent analysis of the findings on risk perception, Kenneth Arrow (1982) concluded that the evidence regarding individual rationality was, at best, quite mixed. A prominent example of apparent irrationality of actual consumer behavior is that consumers, who presumably are risk averse, have failed to purchase heavily subsidized federal flood insurance.[2] In the case of the market for hazardous jobs, which is the focus of this study, Viscusi (1979) found that workers' risk perceptions were positively correlated with the industry risk and that workers who perceived job risks received compensating wage differentials.[3] Nevertheless, workers in high risk jobs displayed behavior consistent with an adaptive response in which workers accept jobs whose risks are not fully understood, learn about these risks based on their on-the-job experiences, and then quit if these experiences are sufficiently unfavorable given the wage for the job.

Although the positive injury rate-quit rate linkage is consistent with an adaptive response, there has been no study that has investigated the dynamics of this relationship. Do workers learn about risks on the job, and does this change in perceptions lead workers to revise their reservation wage rates in the expected manner? More fundamentally, even in the absence of such learning, do workers have subjective risk assessments that generate compensating differentials in the manner that is consistent with studies of risk premiums for hazardous occupations and industries? In this paper we will extend this line of research by analyzing the nature of workers' risk assessments, how workers process information, and how changes in risk perceptions affect their decisions.

Since no existing data sets provide information on the evolution of workers' risk perceptions, we undertook a sample survey in which we ascertained worker responses to labels of potentially hazardous chemicals. We chose this form of information because the chemical industry already has some experience in conveying this information in a manner that workers can understand, thus making it possible to analyze the learning process rather than focusing on the design of the format for the information. In addition, this type of risk information has substantial policy relevance since chemical labeling is the major component of the OSHA hazard communication policy. This $3 billion policy was the most expensive social regulation issued during the first three years of the Reagan Administration.[4]

*Professor and Director, Center for Study of Business Regulation, Fuqua School of Business, Duke University, Durham, NC 27706, and consultant, Green Farms, CT 06436, respectively. This research was supported in part by the Center for the Study of Business Regulation, Duke University. We are indebted to William Evans for a superb computer programming effort, to Henry Feng for compilation of the survey results, and to John Payne and seminar participants at several universities for helpful comments.

[1]A lucid discussion of the relationship of information to economic behavior is provided by Kenneth Arrow (1974).

[2]Howard Kunreuther's (1976) analysis of flood insurance stresses a lack of consumer information as an important factor. A major theme of this study is that workers also are acting within a context of highly imperfect information.

[3]Viscusi (1979) also linked compensating differentials to objective risk measures, yielding comparable wage premiums. Other studies in the compensating differentials literature include Richard Thaler and Sherwin Rosen (1976), Robert Smith (1976), and Charles Brown (1980). Smith (1979) and Viscusi (1983) provide critical surveys of this literature.

[4]The EPA pesticide and toxic chemical regulations also include chemical labeling as a policy option. See Susan Hadden (1983) for a review of the role of labeling policies of federal agencies.

In Section I we discuss the nature of the sample and present the empirical results for the situation before information provision to establish a reference point for subsequent results. These findings are also of interest in their own right because the survey provides extensive risk-related information that included detailed risk-assessment questions and information on whether workers would repeat their job choice. These data will consequently enable us to make a more direct link than in previous research between workers' risk perceptions and labor market outcomes, such as compensating differentials for risk. The effect of the chemical labels on workers' risk perceptions is the subject of Section II. We estimate both the risk level implied by the hazard warning and the informational content relative to the worker's prior beliefs. This evidence is consistent with a Bayesian learning process in which workers retain some influence of their priors and incorporate the new information in the expected manner. Section III's analysis of the effect of risk information on compensating differentials and worker turnover provides the first explicit test of the effect of changes in workers' risk perceptions on labor market performance.

The overall picture that emerges is that workers begin jobs with imperfect information, but there are many rational elements to worker behavior, and the extent of risk-related mismatches of jobs and workers is not rampant. After acquiring risk information, most workers display the capacity to update their probabilistic beliefs in a manner that is broadly consistent with Bayesian analysis. The adaptive responses that emerge suggest that workers are engaged in an ongoing experimentation process in which they learn about the risks posed by their job and quit once the position becomes sufficiently unattractive.

I. The Sample and Baseline Results

Since no existing body of data provides longitudinal information on workers' risk perceptions, we developed a survey to enable us to analyze worker responses to job hazard information. The focus of this section is on the nature of the sample and the empirical results for the situation before workers received risk information. Because of the more comprehensive nature of the risk questions, it is possible to broaden the empirical support for the principal labor market impacts of employment hazards.

The sample consisted of 335 employees in the chemical industry. During the first six months of 1982, the managers responsible for chemical labeling interviewed workers at four plant locations of three major chemical firms. The operations represented included research and development as well as manufacturing. The sample included a broad range of occupational groups exposed to chemicals. Engineers, technicians, chemists, mechanics, researchers, and supervisors were all included. Over half of the sample—187 workers—consisted of workers who were either on hourly pay or were technicians. This group, which we will denote by *BC/TECH*, closely parallels the blue-collar subsample analyzed in Viscusi (1979) and will be the focus of much of the empirical work in this section.

Table 1 summarizes the sample characteristics for the full sample and the *BC/TECH* subsample. The sample characteristics follow the pattern one would expect for a national chemical firm. The average worker age is 39, and the majority of all workers are white males (only 7 percent blacks and 43 percent females). The individuals averaged two years of college education, or years of schooling (*EDUC*) equal to 14. Almost two-thirds of the sample were married with an average of 1.36 children (*KIDS*). Their total work experience (*EXPER*) was 18 years, 8 of which were at the particular firm (*TENURE*). The average annual earnings (*EARNG*) was over $21,000.

The most distinctive characteristic of the sample was the inclusion of a series of risk perception questions. The *DANGER* variable pertains to whether or not the worker's job exposes him to dangerous or unhealthy conditions. The wording of this question parallels that in the University of Michigan *Survey of Working Conditions* (1975) used by Viscusi (1979) and will be used in assessing the comparability of the empirical results. In that study, 52 percent of the blue-collar workers viewed their jobs as dangerous. The

TABLE 1—SAMPLE CHARACTERISTICS: MEANS AND STANDARD DEVIATIONS

Variable	Full Sample	*BC/TECH* Subsample
AGE (in years)	38.8	38.9
	(11.8)	(12.8)
BLACK (0–1 race dummy variable (*dv*))[a]	0.07	0.10
MALE (0–1 sex *dv*)[a]	0.57	0.42
EDUC (years of schooling)	14.44	12.47
	(3.21)	(2.05)
MARRIED (0–1 marital *dv*)[a]	0.64	0.62
KIDS (number of children)	1.36	1.10
	(1.52)	(1.28)
EXPER (years of work experience)	18.38	19.19
	(11.68)	(13.0)
TENURE (years of experience at firm)	8.19	7.15
	(7.22)	(6.41)
EARNG (annual earnings)	$21,120.4	$15,768.6
	($8,332.1)	($3,596.6)
DANGER (0–1 risk *dv*)[a]	.57	0.50
RISK (scaled risk)	0.10	0.09
	(0.06)	(0.07)
HRISK (0–1 high risk *dv*)[a]	0.36	0.35
WPREM (0–1 perceived wage premium *dv*)	0.11	0.10
TAKEA (0–1 repeat job choice *dv*)[a]	0.79	0.77
TAKEB (0–1 repeat job choice *dv*)[a]	0.97	0.96
QUITA (0–1 quit intention *dv*)	0.12	0.12
QUITB (0–1 quit intention *dv*)	0.05	0.05
Sample Size	335	185

[a]Standard deviations for 0–1 dummy variables are omitted since they can be calculated from their fraction m in the sample, where the standard deviation is $(m-m^2)^{.5}$.

results here are quite similar, as 57 percent of the overall sample viewed their jobs as dangerous, with 50 percent of the *BC/TECH* subsample perceiving some risk.

Although the mean *DANGER* levels are not unexpected, the relative riskiness rankings are the opposite of what one might expect since the *BC/TECH* group presumably faces greater risks. Whether or not this is actually the case is not clearcut since the white-collar research chemists may in fact incur greater health risks than, for example, maintenance personnel. The more similar results for the continuous *RISK* variable discussed below suggest, however, that these results may not stem from an actual difference in riskiness. Rather, the *BC/TECH* workers may have a less stringent risk level cutoff for considering whether their jobs are hazardous. Since willingness to accept a risk is negatively related to one's wealth, it is not unexpected that higher-income workers are more likely to regard a job as dangerous, for any given risk level.[5]

Except in the case of one study using the *DANGER* variable, all previous analyses of risk premiums have used objective occupational or industry risk measures. For this paper we developed a variable that would reflect the worker's subjective assessment of the BLS injury and illness frequency rate for his job. From the standpoint of the theoretical foundations of the compensating differential theory, the wage-risk relationship should be driven by such subjective risk perceptions. Aggregative risk variables simply serve as an objective proxy for this variable.

To overcome the difficulties arising from different danger reference points and to provide a continuous risk measure that will make

[5]Educational differences and related differences in ability to perceive risks may also play a role.

possible a detailed analysis of worker learning, we developed a continuous *RISK* variable. We presented to each worker a linear scale, ranging from very safe to dangerous. To provide an objective reference point, an arrow marked the average U.S. private sector injury and illness rate. Each respondent marked on the scale the risk level that he assessed for his job. This variable was then converted into probabilistic terms, that is, scaled between 0 and 1, where risk is on a scale comparable to the BLS annual injury frequency rate. The mean *RISK* levels for the full sample and the *BC/TECH* subsample are comparable to the national average private sector risk probabilities and about 50 percent larger than the recent levels of the chemical industry's injury and illness frequency rate. This discrepancy is not unexpected since BLS statistics primarily capture safety-related accidents and underreport the long-term illnesses from chemical exposures; reported injury rates will understate the actual risk level.

Using the *RISK* responses, we also created a job hazard dummy variable similar to *DANGER* except that the risk threshold reference point was the same for all respondents. The high-risk variable *HRISK* assumed a value of 1 if the worker faced a risk above the U.S. average, and 0 otherwise. A third of the sample viewed their jobs as being high risk, and two-thirds viewed their jobs as being comparatively safe. In conjunction with the earlier *RISK* results, these findings suggest that the chemical industry's relatively good accident record may be a reasonable reflection of most workers' perceptions, but the presence of substantial health risks leads a sizable minority to consider their jobs particularly hazardous.

Since the time of Adam Smith, economists have observed that perceived risks will generate compensating wage differentials since workers will demand extra compensation for jobs that pose extra risk.[6] Table 2 summarizes the risk variable results for equations in which annual earnings (*EARNG*) and its natural logarithm (*LNEARNG*) serve as the dependent variables. Each equation also included an extensive group of variables that typically enter such earnings equations, such as the individual's education and work experience. For the *BC/TECH* subsample, the annual risk premium of \$700–\$800 for *DANGER* was of roughly the same magnitude as the \$900 annual compensation found for the blue-collar subsample in Viscusi (1979) for both *DANGER* and the BLS injury rate.

As with that study, the full sample results were not successful because of an inability to disentangle the wage premiums for risk from the positive overall relationship between job quality and individual income. The change in earnings equations in Section III will not be subject to this difficulty. Restricting the sample to males only eliminates some of the problems arising from failing to control adequately for the omitted variables that determine individual earnings. Male workers' jobs tend to involve more direct handling of chemicals, and the annual risk premiums are considerably larger than for the *BC/TECH* subsample.

Of the three risk variables, *DANGER* yielded the largest annual risk premiums. These were somewhat larger than those for *RISK*, which were about \$100 less. The above-average risk variable *HRISK* led to the smallest annual risk premiums, but the effects were consistently positive and statistically significant (at the 5 percent level, one-tailed test). This pattern may reflect the shortcomings of the *HRISK* variable, which may be a less accurate measure of the underlying job risk, thus leading to a downward bias in its coefficient. The general implications of these findings are less ambiguous. The consistently significant results using the subjective risk variables and the similarity in the *DANGER* and *RISK* premiums to those in earlier studies should bolster one's confidence in the validity of the compensating differential theory.

A closely related issue is whether workers are aware of any risk premiums. Since no previous study had asked workers whether they believed that they received a risk premium, we developed a variable *WPREM* that

[6] See fn. 3 above for a list of several previous risk premium studies.

TABLE 2—SUMMARY OF COMPENSATING DIFFERENTIAL RESULTS[a]

Dependent Variable	Sample	Risk Variable	Risk Coefficient	Average Annual Risk Premium
EARNG	*BC/TECH*	*DANGER*	1577.2 (438.1)	$788.6
LNEARNG	*BC/TECH*	*DANGER*	0.097 (0.029)	$746.5
EARNG	*BC/TECH*	*RISK*	6898.4 (3461.1)	$636.2
LNEARNG	*BC/TECH*	*RISK*	0.479 (0.231)	$665.3
EARNG	*BC/TECH*	*HRISK*	738.4 (465.5)	$258.4
LNEARNG	*BC/TECH*	*HRISK*	0.053 (0.031)	$289.8
EARNG	Full (males)	*DANGER*	2117.5 (775.6)	$1385.7
LNEARNG	Full (males)	*DANGER*	0.124 (0.036)	$1875.3
EARNG	Full (males)	*WPREM*	1583.1 (1179.9)	b
LNEARNG	Full (males)	*WPREM*	.1094 (.0549)	$278.8
EARNG	Full	*DANGER*	169.03 (529.51)	b
LNEARNG	Full	*DANGER*	0.018 (0.025)	b

[a]Each equation also includes the following variables: *AGE*, *BLACK*, *MALE*, *EDUC*, *MARRIED*, *KIDS*, and *EXPER*. The full sample results also include a *BC/TECH* dummy variable. The standard errors are shown in parentheses below the coefficients.
[b]Annual risk premiums are not reported since the coefficients are not statistically significant (at the 5 percent level, one-tailed test).

assumed value of 1 if the worker believed that he received higher pay because of the nature of the chemical industry and 0 otherwise. This variable reflects compensating differentials for working in the chemical industry as opposed to some other industry, not risk premiums per se. Since two-thirds of the sample regarded their jobs as safer than the U.S. average, these incremental premiums should not be large. Only 10 percent of the sample believed they received such a chemical industry premium, and those that did earned an average wage premium of under $300, controlling for other factors (see *LNEARNG* equation, Table 2). As expected, the probability that the worker perceives a risk premium is strongly and positively related to each of the three risk variables, as the logit results in Table 3 indicate.

Since over one-third of the sample believed that they faced above-average risks and only one-tenth acknowledge the existence of relative wage premiums, roughly one-quarter of the sample might appear to behave in a manner that is inconsistent with the standard theory. This need not be the case since workers may, for example, earn some form of economic rent that makes the job attractive despite the absence of a perceived relative risk premium. Moreover, since the overall risk premiums average under $1,000 annually and only $300 for the relative chemical industry differential, many respondents may not have believed that the risk premium they received was sufficiently large to make the chemical industry salary substantially different from what could be earned elsewhere.

Some portion of this group who perceive risks but not relative risk premiums may, however, be mismatched. On a conceptual basis, there clearly is some potential for some

TABLE 3—MAXIMUM-LIKELIHOOD ESTIMATES FOR PERCEIVED RISK PREMIUM AND TURNOVER EQUATION[a]

Dependent Variable	Risk Variable	Coefficient[b]
WPREM	*DANGER*	2.96 (0.75)
WPREM	*RISK*	6.89 (2.78)
WPREM	*HRISK*	0.54 (0.38)
TAKEA	*DANGER*	−1.42 (0.35)
TAKEA	*RISK*	−11.22 (2.32)
TAKEA	*HRISK*	−1.53 (0.30)
QUITA	*DANGER*	1.21 (0.48)
QUITA	*RISK*	6.95 (2.86)
QUITA	*HRISK*	1.55 (0.42)

[a]Other variables entered in each equation include: *AGE*, *BLACK*, *MALE*, *EDUC*, *MARRIED*, *KIDS*, *EXPER* (in *WPREM* equations), *TENURE* (in all except *WPREM* equations), and *EARNG* (in all except *WPREM* equations).

[b]Asymptotic standard errors are shown in parentheses.

labor market mismatches even with rational behavior if workers have some imperfect knowledge of the risks of the job which they continually update as they acquire additional information through their on-the-job experience.[7] Wage premiums for risk will be observed, but workers in high-risk situations will also tend to quit once they have learned about the risks and have decided that the risk compensation is insufficient. Although past empirical work has focused on worker quitting,[8] a related prediction is that if workers were asked to repeat their job choice based on current information, many workers in high-risk jobs would be reluctant to do so. Unlike worker quitting, this job acceptance question is not influenced by transactions costs of job changes, such as seniority rights. This question also avoids the limitations of the relative risk premium question, which may not fully capture the overall desirability of the job.

For the full sample, 79 percent of the sample would decide without hesitation to take the same job (*TAKEA*). The remaining 21 percent would either have some second thoughts or would definitely not take the job. Since 97 percent of all respondents would, at most, "have some second thoughts" (*TAKEB*), only 3 percent of the sample appears to have strong reservations about their positions. The combination of the wage premium estimates and the widespread willingness to repeat the employment decision suggests that job risks are not a major source of worker dissatisfaction. Few workers appear to be seriously mismatched.

One mechanism by which mismatches are remedied is through worker quitting. To analyze the job hazard-quit relationship, we developed quit intention variables utilizing the same phrasing as did the *Survey of Working Conditions* questions analyzed in Viscusi (1979). As shown in that study, this quit intention measure yielded results that were quite similar to those generated by actual quit behavior. One-eighth of the sample was very likely or somewhat likely to "make a genuine effort to find a new job with another employer within the next year" (*QUITA*), but only 5 percent were very likely to do so (*QUITB*). Some worker dissatisfaction is clearly present, but there is not a large proportion of severely dissatisfied workers at the firms in our sample.

The worker's job risk plays an instrumental role in the cases in which mismatches are observed. Table 3 presents the maximum likelihood estimates for the determinants of two job satisfaction measures. In each case, the equations also included a series of variables, such as worker age, that are strongly linked to worker turnover. The probability that the worker would repeat this job decision (*TAKEA*) is negatively related to all perceived risk variables, controlling for

[7]See Viscusi (1979) for a formal presentation of this model.

[8]The job hazard-quit results in Viscusi (1979) are presented for aggregative quit rates, three national samples of panel quit data (Panel Study of Income Dynamics and two National Longitudinal Surveys), and quit intention data from the *Survey of Working Conditions*.

worker earnings and other related factors. A worker who views his job as dangerous (*DANGER*), for example, will have a probability of repeating his initial job choice that is .22 lower than those who do not. Similarly, all of the job risk variables exert a positive influence on *QUITA*, where the quit intention probability will increase from .06 to .19 —or over triple—if the worker views his job as dangerous. Put somewhat differently, the mean effect of the *DANGER* variable accounts for one-half of all quit intentions.

These results are consistent with a model in which the worker's job choice among potentially hazardous jobs is part of an ongoing adaptive process. Workers' reservation wages will increase as their perceived risks rise so that we will observe risk premiums for prior perceived risks and for some risks discovered on the job. Risk that workers learn about but for which they are not compensated sufficiently will generate quits. While the evidence is consistent with this general view, the intermediate learning linkage and the behavioral implications of changes in risk assessments have not yet been examined.

II. Hazard Information and Risk Perceptions

To obtain evidence on this learning process, we carried out the following risk information processing experiment in the second part of the questionnaire. We presented each worker with a hazard warning label for one chemical that was not a current part of his job. Each respondent was told that he would use 100 lb. containers of this substance within the context of his current job operations, but that this chemical would replace the chemicals with which the individual was currently working. The scenario was similar to that in which a worker learns that the chemicals he uses have been mislabelled. We provided workers with "new information" rather than informing them of existing hazards so as to be able to distinguish the role of the hazard warning from a priori knowledge about the job, thus providing a context in which learning could be observed. We then asked each worker how this change would affect his risk perception and other aspects of his behavior. Subsequent changes in risk perceptions consequently do not reflect an inadequacy in workers' prior judgments, but rather how information regarding a newly introduced risk will alter the assessment of the job's implications.

We assigned workers to one of four different labeling groups: sodium bicarbonate (*CARB*), a lachrymator chloroacetophenone (*LAC*), asbestos (*ASB*), and *TNT*. The *CARB* control group was set at a relatively smaller size since the primary focus was on the implications of the three risky substances. Each of these workers was given the information following a standard chemical labeling format. Representative portions of each label are given below:

> SODIUM BICARBONATE. SPILL: Sweep-up, place in an appropriate chemical waste container...

> CHLOROACETOPHENONE. WARNING! LACHRYMATOR—VAPOR AND DUST EXTREMELY IRRITATING. Do not breathe dust or vapor. Wear a self-contained breathing apparatus...

> ASBESTOS. DANGER! CANCER HAZARD. Use with a NIOSH-Mesa approved respirator. Use with approved goggles...

> TNT—(blend of dry Trinitrotoluene). DANGER! HIGH EXPLOSIVES. MUST BE STORED IN ACCORDANCE WITH FEDERAL REGULATIONS. KEEP IN COOL, DRY, WELL VENTILATED, LOCK-UP AREA...

Workers did not proceed with the rest of the questionnaire until they had been able to answer successfully some basic overall questions to test whether they had read the label. The workers appeared to have little difficulty in this regard since they had substantial experience using chemicals labeled in this manner. Although the information provided was not with respect to a specific risk level but for a chemical hazard for which risk assessments will vary, the responses were

TABLE 4—VARIABLE MEANS FOR EACH LABELING GROUP

Risk Variable	Means of Variables with Alternative Labels			
	CARB	*LAC*	*ASB*	*TNT*
RISK	.12	.10	.09	.10
RISK1	.06	.18	.26	.31
HRISK	.42	.38	.29	.40
HRISK1	.07	.83	.95	.98
WBOOST	.03	.48	.71	.82
Risk Premium[a]	0	\$1,919.01	\$2,995.59	\$5,158.31
NOWAGE	0	.02	.11	.17
QUITA	.23	.10	.13	.10
QUITA1	.00	.23	.65	.73
TAKEA	.67	.82	.80	.76
TAKEA1	.90	.58	.11	.07
Sample Size	31	106	102	96

[a] Risk premium is $Y1 - Y$. The figures are conditional upon facing an increased risk and being willing to accept a finite risk premium.

consistent with the general patterns one might expect. Table 4 summarizes the variable means for each labeling group.

Before analyzing the principal economic implications of the learning process, we will first review the general pattern of the responses and their plausibility. Sodium bicarbonate is a very safe substance, and this label leads to a reduction in the *RISK* variable from .12 to .6 for *RISK1*, where the postscript *1* indicates the post-information analogue of the variable. Besides halving the assessed *RISK* level, *CARB* also dramatically reduced the fraction of workers who believe they face above-average risk. Only one respondent raised his *RISK* assessment (from .05 to .06), but since this worker was in a very low-risk job and had a posterior *RISK1* value identical to the *CARB* subsample mean, this behavior cannot be regarded as irrational.

If *CARB* were the only risk posed by the worker's job, one would expect that the workers would assess this risk as being zero. Even when working with a safe substance, there is, however, a residual risk such as the risk of a safety-related job injury from accidents. Since the *RISK1* value of .06 for *CARB* equals the 1980 and 1981 average BLS injury rate for the chemical industry, the results are not out of line with what one might expect once the chemical hazards have been eliminated. In addition, not all workers may have known what sodium bicarbonate is. The label suggests that it is a very safe chemical, but it does not explicitly state that it is risk free.

The lachrymator was the second safest substance in the labeling group. Workers viewed this chemical as more hazardous than their present environment, as the *RISK* level almost doubled, and the fraction of workers who considered themselves in above-average risk jobs increased by .45. Eleven workers did not revise their risk assessments upward after seeing the *LAC* label, but these workers were in very high-risk jobs; their *RISK* level decreased from .19 to .15, which is still above the average pre-information *RISK* value for the sample. Notwithstanding the absence of any assessed increase in risk for this subgroup, one person indicated that he was somewhat likely to look for a new job (*QUITA1*) even though he had not expressed this intention earlier, producing a minor consistency problem.

The asbestos warning led to a more dramatic response. The riskiness of this substance relative to *TNT* is not clearcut because of the deferred nature of asbestos-related cancers. Asbestos is, however, a very potent carcinogen, and it led workers to triple their assessed *RISK* levels, with almost all workers viewing their jobs as above average in riskiness. Somewhat surprisingly, 5 percent of all workers did not view *ASB* jobs as posing above-average risk. Moreover, a substantial group of 26 workers, most of

whom were in very high-risk jobs, did not raise their risk perceptions. The unresponsive group's reservation wage and quit responses (for example, no increase in quits and elimination of all $QUITA = 1$ values) were consistent with their *RISK1* values, so that the *RISK1* variable appears to reflect a more favorable assessment of the job's attractiveness. Such a favorable response is not implausible, particularly for researchers who work with new unregulated carcinogens on a daily basis.

The explosive hazards of *TNT* generated the greatest risk assessment response, as all but two workers now viewed their jobs as above average in risk. Although 11 workers did not raise their *RISK* assessments in response to the warning, these workers were on very hazardous jobs (*RISK* equal to .19), and on average the *TNT* warning lowered their *RISK* value by only .04. There was, however, one seemingly inconsistent respondent who indicated that he was somewhat likely to quit (*QUITA1*) even though he hadn't been earlier, and his assessed *RISK* level had not increased.

As with the earlier results, there is a widespread response to information in the expected direction. The behavior of only a small minority of the workers does not appear consistent with a rational learning process. This result does not, however, imply that workers respond perfectly to new information since the relation between the four labels and actual risk levels is not narrowly defined. Some imprecision is inherent because of differences in individual susceptibility to risk.

To test the empirical implications of the hazard warnings more fully, we will formalize the nature of the learning process. The assumption here is that workers adopt a Bayesian learning approach where their assessed probabilities belong to the *beta* family. This distribution is ideally suited to analyzing independent Bernoulli trials on events such as whether or not one suffers a job accident.[9] We will view the receipt of the new labeling information as equivalent to observing additional Bernoulli trials concerning the riskiness of the job. The implicit assumption is consequently that labels simply serve to augment the risk information available to workers.[10]

The two parameters of the prior distribution are p, the assessed prior probability of an adverse outcome (i.e., *RISK*), and γ, a term that can be regarded as the precision of the prior. After observing m unsuccessful outcomes (for example, accidents) and n successful outcomes, the posterior accident probability $(\gamma p + m)/(\gamma + m + n)$. The term γ is tantamount to the number of trials the worker acts as if he has experienced when forming his prior.

The informational content of each label i likewise depends on two parameters: ξ_i, the precision of the information (i.e., the equivalent number of observations $m + n$ reflected in the information) and s_i, the fraction of these observations that are unfavorable. Whether or not the label raises workers' probability assessments depends on whether s_i exceeds p, and the extent of revision is positively related to the informational content ξ_i, for any given value of s_i. If workers are provided with perfect information and if the labeled chemical is the only risk, the value of ξ_i should be infinite. The labels do not specify the exact chemical risk, so that ξ_i need not be infinite in practice. Moreover, the label only conveys information regarding the risks from direct chemical use so that all accident-related risks and all environmental chemical risks remain. Worker responses consequently will reflect the relative weights workers placed on the prior and posterior information, where these weights will capture both the degree to which the information in the label was credible and the relative role of this risk in the new version of the worker's job.

[9] See John Pratt, Howard Raiffa, and Robert Schlaifer (1975) for more detailed advocacy of the use of *beta* distributions for Bernoulli processes.

[10] If workers do not in fact treat the label as equivalent to additional job experiences but rather "forget" their earlier knowledge, no difficulties are caused provided that the degree of forgetting is determined by the precision of their judgments, not the level of the risk. If the initial risk level were also to affect the weight placed on the label, the empirical estimates would be biased.

The posterior probability p_i of an adverse job outcome after receiving a hazard warning for chemical i is given by

$$(1)\quad p_i^* = \frac{\gamma p + \xi_i s_i}{\gamma + \xi_i} = \frac{\xi_i s_i}{\gamma + \xi_i} + \frac{\gamma p}{\gamma + \xi_i}.$$

The regression equation counterpart of equation (1) for each chemical i is

$$RISK1_i = \alpha_i + \beta_i RISK_i + u_i,$$

where u_i is a random error term and

$$(2)\quad \alpha_i = \frac{\xi_i s_i}{\gamma + \xi_i};\quad \beta_i = \frac{\gamma}{\gamma + \xi_i}.$$

To take into account the bounded nature of the dependent variable, we will also estimate the equations in terms of the log-odds of the probability, or $\ln(RISK/(1-RISK))$. In this case, the parameters α_i and β_i for the linear regression counterpart can be derived from the regression results but are not produced as directly.

The estimated versions of the parameters in equation (2) also can be used to construct two key measures of the information conveyed by the warning. The risk level s_i is given by

$$(3)\quad s_i = -\alpha_i/(\beta_i - 1),$$

which can be verified using equation (2). Similarly, the informational content of the warning relative to the prior, Ψ_i, is given by

$$(4)\quad \Psi_i \equiv \xi_i/\gamma = (1/\beta_i) - 1.$$

Higher values of Ψ_i imply greater informativeness of the label compared to the worker's initial judgments.

To the extent that workers' risk responses reflect not only changes in the probability of an adverse outcome but also changes in their severity, one must modify the formulas above. Let V_i be the severity (i.e., monetary equivalent) of the health impact posed by the hazard warning relative to that posed by the average U.S. job injury, which serves as the metric for the analysis. If the $RISK1$ responses reflect changes both in the probability of an accident and its severity, equation (3) becomes

$$(5)\quad s_i V_i = -\alpha_i/(\beta_i - 1),$$

and the formulation and interpretation of equation (4) remains unaltered.[11] Although the discussion below will be in probabilistic terms and will not include V_i explicitly, it should be noted that these risks are severity weighted.

Table 5 summarizes the regression results and the parameters calculated from them. Overall, the linear variant of the equation provided a better fit than the log-odds formulation. The coefficients α_i and β_i reflect the nature of the learning process. In the case in which workers' judgments are not affected by the hazard warning and are solely dependent on their prior value of $RISK$, α_i will equal 0 and β_i will equal 1. At the other extreme in which the hazard information is dominant, β_i will equal 0 and α_i will be positive. The regression results were between these two extremes. In all cases the label provided a substantial input, and in two cases the prior continued to play a significant role. These results are broadly consistent with a Bayesian learning model.

In the case of $CARB$, the label lowered the $RISK$ assessment but did not eliminate the role of the prior, as both α_i and β_i were statistically significant in the linear case where the relation to equation (2) is direct. The risk level s_1 implied by $CARB$ is .04, or under half of the worker's prior $RISK$ level, and Ψ_1 implies that the relative precision of

[11]More specifically, let V_0 be the original accident severity and V_i be the severity of the postwarning accident. Suppose that the components of $RISK1$ represent a weighted average of these risks and that they take the form

$$RISK1 = (\gamma p/(\gamma + \xi_i))V_0 + (\xi_i s_i/(\gamma + \xi_i))V_i.$$

If we set V_0 equal to 1 (no loss of generality), the values of α_i and β_i are given by

$$\alpha_i = \xi_i s_i V_i/(\gamma + \xi_i) \quad \text{and} \quad \beta_i = \gamma/(\gamma + \xi_i).$$

The severity-weighted results in the text follow directly.

TABLE 5—RISK PERCEPTION AFTER INFORMATION REGRESSION RESULTS

	CARB		*LAC*		*ASB*		*TNT*	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Constant	0.030	−3.58	0.14	−2.05	0.25	−1.23	0.31	−0.86
	(0.014)	(0.46)	(0.01)	(0.13)	(0.02)	(0.11)	(0.02)	(0.09)
RISK	0.21	3.23	0.44	3.76	0.14	1.36	0.03	0.11
	(0.10)	(3.56)	(0.10)	(1.12)	(0.14)	(1.03)	(0.13)	(0.76)
R^2	.12	.03	.14	.10	.08	.02	.01	.01
s_i	.038	.042	.239	.274	.289	.325	.317	.315
Ψ_i	3.72	4.98	1.29	0.83	6.43	2.80	31.36	40.67

Note: All cols. (1) are *RISK1* (linear); all cols. (2) are *RISK1* (log-odds).

the hazard warning was 4–5 times that of the prior.

Since the very safe properties of sodium bicarbonate are reasonably well known, one might have expected that *CARB* would result in a larger relative precision estimate and a lower s_1 than was observed. A possible explanation is that workers did not place an infinite weight on a chemical exposure with near zero risk because of the residual risks of the job. These workers will continue to be exposed to a variety of airborne carcinogens and safety-related risks that will be reflected in the posterior *RISK1* values. As the risks captured by the label approach zero, the nonzero risk components of the worker's job become more instrumental since they dominate the role of the label.

The *CARB* label was, however, much more powerful than the *LAC* warning. This label led to the greatest retention of workers' prior beliefs, as the *RISK* coefficients are the largest of any of the regressions. A small impact was not a consequence of any close similarity in the hazard probabilities of *LAC* and *RISK*, since s_2 is over double the prior value of *RISK*. The limited nature of the effect derives from the lower relative precision Ψ_2 of this warning, which had roughly the same informational content as did workers' prior beliefs.

Warnings for the severe risks of *ASB* and *TNT* are so powerful that the prior *RISK* variable plays an insignificant role; only the constant terms enter. The risk levels s_3 and s_4 are somewhat higher than for *LAC*, but the major difference is the precision of the information. Asbestos warnings have roughly the same relative precision as *LAC*, but *TNT* has especially large informational content, roughly 30–40 times that of the prior. Since *TNT* poses well known explosive risks, this result is not unexpected.

Overall, the risk levels s_i implied by *LAC*, *ASB*, and *TNT* were not too dissimilar. The greatest difference was the relative precision associated with these warnings. The impact of a hazard warning does not hinge solely on the implied risk level. In this instance, the informational content of the label proved to be more instrumental in altering workers' probabilistic judgments. To be effective, hazard warnings must convey information in a convincing manner. Otherwise, the weight individuals place on their prior beliefs will dominate in the formation of workers' risk judgments.

III. The Effect of Learning on Worker Behavior

The change in the risk perceptions resulting from the hazard warnings in turn will affect worker behavior if workers make sequential decisions in an optimal manner. The data in the bottom portion of Table 4 summarize the wage and turnover effects, which reflect similar patterns of influence. After reviewing the general nature of these responses, we will use these responses to test the key hypotheses regarding rational worker behavior.

The demand for risk premiums is positively related to the change in the risk, as one would expect. The fraction of workers who indicate that they would require a higher salary to be willing to work with the new

chemical (*WBOOST*) is about three-fourths for *ASB* and *TNT*. As noted above, there are some workers with very high initial risk assessments that were not increased as a result of the label so that not all workers will desire extra compensation. The amount of extra compensation demanded ranges from $2,000 for *LAC* to over $5,000 for *TNT*. Workers need no risk premium to work with *CARB*. (Indeed, they should be willing to take a pay cut, but the survey did not address this possibility.) The premium estimates are only for those workers willing to remain on the job in return for extra pay. Some workers, particularly for *TNT* and *ASB*, were not willing to state an acceptable reservation wage (*NOWAGE*). Whether these 29 nonrespondents were unwilling to accept any finite risk premium or simply believed that no adequate risk premium was feasible is unclear.

The effect on worker turnover was particularly dramatic since the experiment altered the risk but did not alter the wage rate. These risks consequently will produce a more dramatic worker response than in a market context where there would be some adjustment in the wage level. The *QUITA1* and *TAKEA1* questions pertained to the attractiveness of the current job, varying only the risk. In the case of *CARB*, there was a 23 percent drop in quit intentions to zero, and an equal increase in the percentage of workers who would repeat their job choice. The lachrymator produced a 13 percent increase in quit intentions and a 24 percent drop in workers willing to repeat their job choice. The strongest effects were for *ASB* and *TNT*, which would lead the majority of workers to quit and almost all workers to be unwilling to repeat their job choice.

An instructive check on the validity of these responses is to analyze whether the behavioral relationship governing the risk premium and quit decisions parallel those in the pre-information situation. Such an analysis will also make possible an explicit test of the impact of the risk s_i implied by the label and its relative precision Ψ_i. Higher implied risks s_i clearly should make the job less attractive. The relative precision of one's risk assessment will also increase workers' reservation wage since, as shown in Viscusi (1979), the value of a risky job is negatively related to the precision of one's risk judgments. Jobs associated with looser probabilistic judgments are more attractive since they offer greater potential gains from experimentation. Workers can terminate uncertain jobs if their learning is unfavorable and reap the high expected rewards from jobs associated with favorable on-the-job experiences. This asymmetry generates a predilection for loose priors. This aspect of adaptive behavior is the most distinctive prediction of the model, but it has never been the subject of an explicit empirical test.

To analyze the effect of the hazard warnings on the level of compensating differentials, we first need some additional notation. Let Y represent initial worker income, X be a vector of all nonrisk variables for that job, Z be the unmeasured effects specific to the job-worker match, and u be the error term. The compensating differential results in Section I focused on an equation of the form

$$(6)\quad Y = \beta X + \beta^{*} RISK + \beta^{**}\gamma + Z + u.$$

Since γ and Z were omitted from the model, the estimated coefficients were subject to omitted variables bias.

The situation following information (denoted by postscript *1*) can be modelled similarly, where

$$(7)\quad Y1 = \beta X + \beta^{*} RISK1 + \beta^{**}(\gamma + \xi) + Z + u1.$$

Subtracting equation 6 from equation 7 yields

$$(8)\quad Y1 - Y = \beta^{*}\Delta RISK + \beta^{**}\xi + u1 - u,$$

where $\Delta RISK$ is $RISK1 - RISK$. Equation (8) will yield consistent estimates of the coefficients in this fixed effects model as the sample size $N \to \infty$ if there is sufficient variation in $\Delta RISK$ and ξ.[12] It should be noted that we do not have information on ξ_i but

[12] Use of the fixed effects model in compensating differentials studies is not unprecedented. See Brown and, more generally, see Gary Chamberlain (1982).

TABLE 6—POST-INFORMATION EARNINGS AND QUIT EQUATIONS[a]

Dependent Variable	*RISK* or Δ*RISK*	s	Ψ	$R^2/-2$ Log Likelihood
EARNG	9934.5	6784.2[b]	52.0[b]	.24
	(5468.6)	(3342.3)	(32.2)	
Δ*EARNG*	12435.3	–	65.56[b]	.17
	(2681.9)		(20.10)	
EARNG	9838.3	6602.3[c]	41.6[c]	.24
	(5471.3)	(2684.2)	(19.6)	
Δ*EARNG*	12777.5	–	46.53[c]	.17
	(2640.3)		(14.17)	
LNEARNG	.627	.456[b]	.0021[b]	.28
	(.303)	(.185)	(.0018)	
Δ*LNEARNG*	.633	–	.0027[b]	.31
	(.087)		(.0007)	
LNEARNG	.622	.424[c]	.0018[c]	.28
	(.303)	(.149)	(.0011)	
Δ*LNEARNG*	.651	–	.0019[c]	.31
	(.086)		(.0005)	
QUITA	−1.05	5.95[b]	.027[b]	381.3
	(2.03)	(1.43)	(.013)	
Δ*QUITA*	20.4	–	.011[b]	63.3
	(4.3)		(.029)	
QUITA	−1.07	5.75[c]	.021[c]	384.6
	(2.02)	(1.17)	(.008)	
Δ*QUITA*	20.6	–	.002[c]	63.5
	(4.3)		(.020)	

[a]All cross-sectional equations include other explanatory variables as in Tables 2 and 3.
[b]The Ψ variable is based on the linear regression estimates reported in Table 5.
[c]The Ψ variable is based on the log-odds regression estimates reported in Table 5.

on Ψ_i for each labeling group, which is ξ_i/γ. Workers, however, will differ in the precision of their priors, so that γ will be a random variable. Since the workers were assigned randomly to each labeling group, the precision variable should be subject to random measurement error, biasing the β^{**} coefficient downward.

Table 6 reports the earnings equations both in the first difference form (i.e., Δ*EARNG*, Δ*LNEARNG*) and in the cross-sectional form for the post-information case, where the *RISK* variable is of the same form as the dependent variable (Δ*RISK*). Since the first differencing eliminates the biases from omitted fixed effects, the change in earnings equations will be estimated for the full sample, while the cross-sectional results will focus on the *BC/TECH* subsample as before. In the case of the post-information cross section, we included both *RISK* and s rather than *RISK1* in order to estimate explicitly the role of the risk implied by the warning. The results reflect a consistent pattern of premiums for prior risks and risks communicated through the label. Similarly, labels associated with high relative precision Ψ generate additional premiums, as predicted.

The consistency of worker behavior with the earlier results is more difficult to ascertain since premiums per unit of risk should be larger since individuals will demand higher rates of compensation if placed in a highly risky job that is not consistent with their preferences. Whereas the initially perceived risks are the result of a voluntary self-selection process, the post-information risks are not, and serious mismatches may occur. Higher desired premiums per unit of risk consequently should be observed.

The magnitude of the post-information wage-risk tradeoff bears out this pattern. In the case of the linear specifications, for example, the *RISK* and s coefficients average

about one-fifth higher than in Table 2, while in the first difference form $\Delta RISK$ commands premiums three-fourths larger. A greater response is observed in the first differencing case because the additional desired premiums per unit of risk for the added hazard will be averaged only across the extra risks, whereas the post-information cross section obtains an average unit risk premium for the entire risk level. In addition, about one-third of the discrepency arises because the first differencing results focus on the full sample, which is wealthier than the *BC/TECH* subsample used in the cross sectional results. These workers consequently demand a larger premium per unit risk.[13]

To analyze the change in workers' quit decisions, we can formulate a post-information cross section and an analogue of the fixed effects model for discrete variables.[14] The post-information quit intentions in the cross-sectional results are driven exclusively by the implied risk and precision of the hazard, each of which has the expected positive effect. The most dramatic difference with the earlier results is in the $\Delta RISK$ coefficients in the first difference equations, which are almost three times larger than in the preinformation results in Table 3. Such a dramatic increase is not implausible since quits arising in the market are in response to a pay-risk package mix that the worker initially accepted. Here workers are responding to often dramatic changes in their job's attractiveness so that the intensity of the response should increase. The $\Delta QUITA$ equations do not, however, lead to significant coefficients for Ψ, a result that may be due to the drop in sample size down to 161 as a consequence of the statistical estimation procedure that has been used.

IV. Conclusion

The focus of this analysis has been on an adaptive framework in which individuals do not have perfect job risk information, but instead continually revise their risk judgments in Bayesian fashion and then switch jobs once these judgments become too unfavorable. This theory is an extension of the standard compensating differential analysis rather than an incompatible theory. Workers' initial perceptions of risk led to compensating differentials and also generated intentions to quit and regret over having accepted the job initially. The evidence of risk-related job mismatches is consistent with a model of job experimentation and would not occur in a perfect information version of the compensating differential model. The extent of these mismatches does not, however, appear to be great, so that for this sample the market appears to operate reasonably effectively.

After being given a hazard warning for use of a new chemical in their job, workers revised their risk assessments in the expected directions, but retained some influence of their prior for hazard warnings with low informational content. Although the risk level implied by the label was of consequence, differences in informational content appeared to be more influential in governing one's posterior risk assessment. This learning in turn generated a demand for risk premiums and incentives to quit, as predicted. Both the change in the level of the risk and changes in the precision of workers' judgments were of consequence, as the adaptive model predicts. Although the change in the risk level had a more consistent direct effect on behavior than did the relative precision of the hazard warning, the precision also has an indirect influence through its powerful impact on the posterior risk assessment.

The pivotal influence of the informational content of the chemical label has broad ramifications for the design of effective risk information strategies. Past informational campaigns such as those intended to encour-

[13] Estimation of the *EARNG* and *LNEARNG* equations for the *BC/TECH* subsample yielded annual risk premiums about $1,000 less than for the full sample.

[14] Using the procedure developed by Chamberlain (1980), we will restrict the sample to those individuals who altered their quit decisions since sample observations involving the same quit responses provide no useful information for the estimation. Those (0,1) responses who would quit after the warning but not before (primarily from *LAC*, *TNT*, and *ASB* groups) constitute one of the binary outcomes and the (1,0) responses (primarily from *CARB*) constitute the other outcome. The explanatory variables are the first differences of the variables included in the pre-information equation so that only the risk-related variables remain.

age seatbelt use and deter cigarette smoking have generated disappointing results. The primary purpose of these efforts is that of exhortation rather than providing consumers with information that they did not already possess. The lack of a major consumer response should not be unexpected since the informational content of these warnings was low. The results in this study indicate that risk information programs will be most effective when they do not simply convey the risk level, but they also provide individuals with new information in a convincing manner.

Most workers behaved as expected, but there was a small minority of alarmist responses as well as some inertia and inconsistencies. Moreover, while the empirical evidence constitutes the most refined test of the Bayesian learning model of adaptive job choice, observed consistency with the principal predictions of the theory does not necessarily imply full rationality. Nevertheless, there is strong evidence of a systematic worker response that is quite different from the polar extremes of optimal decisions with perfect prior information and random decisions by irrational workers.

REFERENCES

Arrow, Kenneth, "Limited Knowledge and Economic Analysis," *American Economic Review*, March 1974, *64*, 1–10.

______, "Risk Perception in Psychology and Economics," *Economic Inquiry*, January 1982, *20*, 1–9.

Brown, Charles, "Equalizing Differences in the Labor Market," *Quarterly Journal of Economics*, February 1980, *94*, 113–34.

Chamberlain, Gary, "Analysis of Covariance with Qualitative Data," *Review of Economic Studies*, January 1980, *47*, 225–38.

______, "Panel Data," Working Paper No. 913, National Bureau of Economic Research, 1982.

Hadden, Susan, *Read the Label: Reducing Risk by Providing Information*, draft book manuscript, University of Texas-Austin, July 1983.

Kunreuther, Howard, "Limited Knowledge and Insurance Protection," *Public Policy*, Spring 1976, *24*, 227–61.

Pratt, John, Raiffa, Howard and Schlaifer, Robert, *Introduction to Statistical Decision Theory*, New York: McGraw-Hill, 1975.

Smith, Adam, *Wealth of Nations*, New York: Modern Library, 1937.

Smith, Robert S., "Compensating Differentials and Public Policy: A Review," *Industrial and Labor Relations Review*, Spring 1979, *32*, 339–52.

______, *The Occupational Safety and Health Act: Its Goals and Achievements*, Washington: American Enterprise Institute, 1976.

Thaler, Richard and Rosen, Sherwin, "The Value of Saving a Life: Evidence from the Labor Market," in N. Terleckyj, ed., *Household Production and Consumption*, NBER Studies in Income and Wealth No. 40, New York: Columbia University Press, 1976.

Viscusi, W. Kip, *Employment Hazards: An Investigation of Market Performance*, Cambridge: Harvard University Press, 1979.

______, *Risk by Choice: Regulating Health and Safety in the Workplace*, Cambridge: Harvard University Press, 1983.

Institute of Social Research, *Survey of Working Conditions*, SRC Study No. 45369, Ann Arbor: University of Michigan Social Science Archive, 1975.

Migration and Interregional Employment Redistribution in the United States

By Michael J. Greenwood and Gary L. Hunt*

During the 1970's the South and West census regions accounted for 89.9 percent of incremental U.S. population and 72.5 percent of incremental nonagricultural employment. These figures are sharply higher than corresponding figures for any other decade in U.S. history. For example, during the 1930's the South and West gained 65.4 percent of incremental national population, which was the highest share previously recorded. The immediate cause of the interregional population shifts is fairly heavy net out-migration from the Northeast and North Central regions to the South and West, combined with birth rates that are near historical low levels. Interactive relationships that cause employment growth and migration to reinforce one another are probably operating, but these relationships are not well understood. They are the focus of this paper.

Using a new and unique set of time-series data on migration and employment, we develop and estimate a time-series model of migration and employment growth for each of 171 U.S. regions. Our primary emphasis in this paper is on the 57 major metropolitan areas.[1] We demonstrate that incremental employment opportunities are differentially attractive to migrants if the opportunities occur in southern and western areas. Consequently, national employment expansion benefits southern and western areas in an indirect way. Whereas the various areas, without regard to their location, appear to capture their approximate share of direct incremental national employment, the South and West ultimately gain a somewhat larger share of incremental employment. The reason is that the newly available employment opportunities in southern and western areas attract migrants to these areas and away from those located in the Northeast and North Central regions.

Historically, migration to cities has probably been a self-reinforcing and cumulative phenomenon. Many factors may underlie such a relationship. Among these are: 1) the skills, inventiveness, and innovativeness of the migrants themselves, who may possess differential endowments of human capital (in the form of education, accumulated skills, or entrepreneurial talent, for example) relative to the population of the sending or receiving areas. 2) Apart from their human capital, migrants may own physical and financial capital that they bring with them. 3) Migrants may possess sources of income other than their labor services. In 1978, over one-third of U.S. personal income was from sources other than wage and salary income. 4) As stressed by George Borts and Jerome Stein (1964), migrants may cause increased investment (for example, social infrastructure, housing) in receiving localities. 5) Migrants may influence the price and profitability of locally provided goods and services due to the changed demand they may cause for such goods and services. 6) Migrants may contribute to the growth of markets and to the achievement of scale and agglomeration economies, as suggested by Gerald Goldstein and Leon Moses (1973).

Note that each of the above factors would cause a shift of the local labor demand schedule, as opposed to a movement along this schedule caused by a shift of the local

*Professor, Department of Economics, University of Colorado, Boulder, CO 80309, and Economist, Research Department, Federal Reserve Bank of Chicago, IL 60690. We are grateful to Philip Graves and to two referees for a number of helpful suggestions. The early stages of this research were supported by a grant from the Economic Development Administration, U.S. Department of Commerce.

[1]The major reason for our focus on the 57 largest areas, or top one-third of the sample in terms of size, is convenience of exposition. These 57 areas held an average of 77.1 percent of national employment over the study period. In every essential respect, the qualitative results are identical when all 171 areas are studied.

labor supply schedule. Each of the above factors could have the result that migration during any given period causes an increase in employment during that or subsequent periods that exceeds the migrants' own direct contribution to area employment. The resulting excess labor demand could in turn be satisfied by still more migration and/or by increased labor force participation of indigenous residents. Outward migration from many large U.S. cities and broader metropolitan areas, such as has been observed over the past decade, may have just the opposite consequences.

This study addresses two questions that are pertinent to the issue of the interrelationship between migration and urban employment growth, namely, how many employed net migrants are directly attracted by a given number of incremental jobs, and how many extra jobs result directly from the in-movement of a given number of employed persons? Each of these questions could, of course, be phrased in terms of employment decline and out-migration.

The only prior study that yields immediate insight into these questions is that of Richard Muth (1971), who develops a two-equation model, with one equation for net migration and one for employment change.[2] His results, which focus on the 1950–60 period, indicate that net in-migration causes a direct increase in employment that is approximately proportional to the increase in the labor force caused by migration. Hence, employment rises by the employed migrants' contribution to employment, not by more and not by less. Moreover, Muth finds that employment growth also attracts migration, but that the magnitude of this relationship is somewhat less than one to one; rather, it is more like 0.6 or 0.7 to one. That is, every 10 incremental jobs are filled by between 6 and 7 net in-migrants.[3]

An elasticity of migration with respect to employment change of 0.6 or 0.7 implies that roughly two-thirds of the incremental jobs that occurred between 1950 and 1960 in large urbanized areas were filled by migrants, with the remaining one-third filled by the previously unemployed and by new labor force entrants. The longer the period over which migration is defined, the larger the expected value of this coefficient. Moreover, during a period when labor force entry is high, due either to demographic changes or to an increased rate of labor force participation, relatively few migrants should absorb any given number of incremental jobs because of increased competition from indigenous residents. The period after roughly 1960, which defined the endpoint of Muth's empirical investigation, was one of extremely rapid labor force expansion due both to the aging of the baby-boom cohort and to increased labor force participation rates of women. No alternative estimates of these important relationships have been forthcoming, however.

The present study differs from Muth's in a number of ways. First, whereas Muth developed estimates of labor force migration by using cohort survival techniques and by applying labor force participation rates to various age-specific migration groups, the data used here include a direct measure of employment migration.[4] Second, while Muth used cross-sectional data, time-series data applied to each of several large areas are employed in this study. Third, except as regards the mutual interaction between employment change and net migration, the models that underlie the two studies are considerably different. Muth implicitly assumes that another job is equally attractive to migrants no matter where the job is located.[5]

[2]Also see Muth (1968).

[3]Muth appropriately stresses the idea that such structural coefficients can give a misleading impression of the overall impacts of an exogenous shift of either the migration or the employment equation. When the interactions inherent in his model are taken into account, an exogenous increase of one job induces two employed net in-migrants, with the consequence that total employment rises by three jobs.

[4]See Warren Mazek and John Chang (1972) for comments on Muth's measure of migration. Note also that because it is a measure of migration obtained as a residual, Muth's measure contains an unknown degree of error.

[5]This statement should be qualified somewhat in that Muth does present separate estimates for the Northeast, for the Northeast and South, for the Northeast, South, and West, for large, for medium, and for small cities.

This study shows an unmistakable tendency for the migrant-attractive power of another job to vary from place to place.

I. The Data

Since the data used in this study are unique and importantly determined the form of the estimated model, let us briefly consider the nature of the data. Until a short time ago, comprehensive time-series data on migration in the United States were nonexistent. Consequently, interregional migration in the United States has almost exclusively been analyzed within a cross-sectional framework. However, recently developed times-series data on migration, employment, and earnings allow the estimation of temporal models of migration, which are the focus of this study. The data utilized here are unique in that they are the only available U.S. data that yield an annual measure of migration for a reasonably long period of time, and that also have nationally inclusive spatial detail for a reasonably large number of areas.

The migration data are derived from a 1 percent sample of all persons employed in Social Security-covered jobs during each year from 1958 to 1975. Although the One-Percent Social Security Continuous Work History Sample (*CWHS*) has been available for many years, the data contained in the file were only recently expressed on an annual basis. The spatial unit employed in this study is the Bureau of Economic Analysis (BEA) Economic Area. The coterminous United States is divided into 171 such areas, with each area presumably delineating a labor market. Hence, 171 observations are available for each of 18 years. The major focus of this study is on the 57 areas that had the largest average employment level over the 19 annual data points (from the end of 1957 to the end of 1975).

Migration refers to a change during the course of a year in the BEA Economic Area in which an individual's place of work is located, as reported by his employer. All migration, therefore, relates to persons employed both before and after moving. Employment and earnings data are also from the *CWHS* and are consistent with the migration data. Change in area employment during a given year is the difference between an area's Social Security employment at the end of the year and the area's Social Security employment at the beginning of the year.[6] Average earnings levels in the various areas also refer strictly to earnings from Social Security-covered employment.

II. The Model

Apart from data contained in the *CWHS*, little time-series information is available on BEA Economic Areas, which necessitates the use of a fairly simple model. Moreover, degrees of freedom constraints restrict the number of variables that can reasonably be employed in a given regression. The model estimated in this study consists of a simultaneous system made up of five equations, three of which are behavioral, one of which is a nonlinear identity, and one of which is a linear identity. The behavioral equations are for net in-migration to area i during period t, for employment change in i during t, and for change in average real annual earnings in i during t. Specifically, the equations for a representative area are of the following forms:

$$(1)\quad NM_t = \alpha_0 + (\alpha_1 + \alpha_2)\Delta\hat{E}_t + \alpha_2\Delta\hat{E}US_t' + \alpha_3 RW_{t-1} + \alpha_4 NM_{t-1},$$

$$(2)\quad \Delta E_t = \beta_0 + \beta_1\hat{N}M_t + \beta_2\Delta EUS_t + \beta_3\Delta\hat{Y}_t + \beta_4\Delta E_{t-1},$$

$$(3)\quad \Delta W_t = \gamma_0 + \gamma_1\hat{N}M_t + \gamma_2\Delta WUS_t + \gamma_3\Delta\hat{Y}_t + \gamma_4\Delta W_{t-1},$$

$$(4)\quad \Delta Y_t \equiv (W_{t-1}\cdot\Delta E_t) + (E_{t-1}\cdot\Delta W_t) + (\Delta E_t\cdot\Delta W_t),$$

$$(5)\quad \Delta EUS_t' \equiv \Delta EUS_t - \Delta E_t,$$

[6] Employment in agriculture, forestry, and fishing has been removed from the data because of measurement problems inherent in Social Security employment for this industry. This source of employment accounted for an average of only 2.2 percent of Social Security-covered employment.

where NM_t = absolute net in-migration of employed persons to area i during period t, in hundreds;

ΔE_t = absolute change in employment in i during t, in hundreds;

ΔEUS_t = absolute change in U.S. employment during t, in hundreds;

$\Delta EUS_t'$ = absolute change in U.S. employment in all areas except i during t, in hundreds;

E_{t-1} = absolute employment in i at the beginning of year t, in hundreds;

ΔW_t = absolute change in average real annual earnings in i during t, where for each area annual earnings are deflated by the national consumer price index;

ΔWUS_t = absolute change in average real annual earnings in the United States during t;

RW_{t-1} = average annual earnings in area i at the beginning of year t relative to U.S. average annual earnings at the beginning of t;

W_{t-1} = average annual real earnings per employed worker in area i at the beginning of year t; and

ΔY_t = absolute change in total real annual earnings in i, which serves as our measure of change in aggregate local demand.

A basic simultaneity inherent in the model, and the one on which this paper focuses, runs between net in-migration to area i during year t and employment change in i during t. The model is structured in such a way that the coefficient (α_1) on ΔE_t in the net migration equation indicates the number of net in-migrants attracted to area i by one additional job. This coefficient is expected to take a value such that $0 \leq \alpha_1 \leq 1$. The range of values indicated here reflects the extreme situations in which local residents absorb all incremental jobs ($\alpha_1 = 0$) and in which migrants absorb all incremental jobs ($\alpha_1 = 1$).[7]

We saw previously that, based upon his estimate that an employed migrant directly results in exactly one more job, as well as on other evidence, Muth infers a perfectly elastic demand curve for local labor. On a priori grounds, however, a number of factors can be identified that allow for the possibility that local demand schedules, rather than being perfectly elastic, actually have negative slope. For example, negatively sloped factor demand curves are easily derived from Cobb-Douglas or *CES* production functions, and most areas are probably not perfectly competitive in their export markets. Shipping costs, which rise with distance, can partially protect local firms from competition. Furthermore, many areas specialize in the production of products that are not sold in competitive markets (for example, Pittsburgh in steel, Detroit in automobiles, and Seattle in airplanes). As suggested above, migration may cause local labor demand as well as local labor supply schedules to shift. Moreover, to the extent that net in-migration results in the achievement of agglomeration economies, local labor demand schedules would not only shift out, but would also flatten. Such flattening could in turn contribute to Muth's finding of a one-to-one relationship between migration and employment change.

The coefficient (β_1) on the net-migration variable in the employment-change equation indicates the number of (direct) jobs that result in area i during period t from one more net in-migrant. While this coefficient is

[7]This point should be qualified in that the Social Security measure of employment change is made up of three components, namely, change in employment due to net migration of persons employed in Social Security-covered jobs both before and after moving, change in employment due to an excess (deficit) of entrants to the Social Security system over exits from the system, and change in employment due to an excess (deficit) of entrants from the military and other sources over exits to the military and other sources. The latter two sources of employment change could include actual migrants who were not counted as such because they were not holding jobs in Social Security-covered employment at either the beginning or the end of the period over which migration was measured. Since no means are available to allow the identification of such individuals, the latter two categories of employment change are referred to as local or indigenous. To the extent that unidentified migrants are included in the measure of indigenous employment change used here, the migrant attractive power of another job is greater than that reported, whereas the number of jobs resulting from another migrant is less. Note that this statement also holds with respect to unemployed migrants who later find a Social-Security covered job.

clearly expected to be positive because net in-migration causes both labor supply and labor demand to increase, whereas net out-migration causes both to decrease, we do not know whether net in-migration results in increased or decreased employment among an area's indigenous residents. Either outcome is possible, with the ultimate consequence dependent upon the relative magnitudes of labor demand and supply shifts, the values of local demand and supply elasticities, and the responsiveness of local wages to the changes that occur. Since a measure of employment migration is utilized in this study, the migrant himself is included in the incremental employment. Hence, $\beta_1 > 1$ implies an increase in employment of indigenous residents (equal to $\beta_1 - 1$), whereas $0 < \beta_1 < 1$ implies a decrease in employment of indigenous residents, or substitution of migrant for indigenous labor. If $\beta_1 = 1$, no change occurs in employment of indigenous residents, and area employment rises by exactly the amount of the migrant's contribution to employment. This latter possibility reflects Muth's earlier finding (1971).

For each area the coefficient (γ_1) on the net-migration variable in the wage equation indicates the magnitude by which local real wages change due to migration. This coefficient could be positive, negative, or zero, depending upon whether the labor demand shift due to migration to i dominates the labor supply shift due to migration to i ($\gamma_1 > 0$), whether the supply shift dominates the demand shift ($\gamma_1 < 0$), or whether the two shifts offset each other ($\gamma_1 = 0$). Although specific labor supply and demand relationships are not estimated for each local area, the results associated with migration's impact on employment (β_1), taken in combination with those associated with migration's impact on wages (γ_1), should allow inferences concerning plausible configurations of such demand and supply relationships and their relative shifts. For example, for a given area, $\beta_1 > 1$ and $\gamma_1 > 0$ could be the result of a positively sloped labor supply schedule, a negatively sloped labor demand schedule, and a demand shift resulting from migration that dominates the supply shift. The same configuration of demand and supply schedules, but with supply shift dominating demand shift, would result in $\beta_1 < 1$ and $\gamma_1 < 0$.[8]

Inclusion of the variables for U.S. employment change in the equation for local employment change is a crude method of tying the temporal performance of employment in each region to the temporal performance of the national economy. Each area sells some fraction of its output in nonlocal markets, and thus some fraction of each area's employment can be directly tied to national economic conditions. In turn, local employment change is tied to employment change in the area's export markets. Thus, a given change in national employment will affect employment in each region, and this change in national employment will be exhaustively shared among the various regions. The coefficient (β_2) on ΔEUS_t indicates the average share of an incremental national job that directly accrues to area i. Other things being equal, a given change in national employment should cause each region to experience an increase in employment, as indicated by β_2.

In turn, again other things being equal, this incremental local employment should attract migrants, as given by α_1. Not only will employment grow in region i, however, it will also grow in every other region, with the consequence that out-migration from these other regions to area i will be discouraged. Thus, α_2, which is expected to take a negative sign, should reflect the amount by which migration to i will decrease due to the increase in national employment that accrues to other regions. Since α_2 has been estimated as the coefficient on ΔEUS and since $\Delta EUS' = \Delta EUS - \Delta E$, where $\Delta EUS'$ is employment change in all areas except for i, the direct effect of employment change in i (α_1) is partially offset by α_2. Note that the "full" direct effect of an additional job in i is thus

[8]Another, though less likely possibility is for the local labor supply curve to be backward bending. In this case, if the demand shift were to dominate the supply shift, local wages would rise ($\gamma_1 > 0$), but employment would rise by less than the migrants' contribution to it ($\beta_1 < 1$). If the supply shift were to dominate the demand shift, then $\gamma_1 < 0$ and $\beta_1 > 1$.

given by $(\alpha_1 + \alpha_2)$. Note also that whereas we treat national incremental employment as exogenously given, we treat the distribution of this employment as endogenous.

The ideal manner in which to model this phenomenon would be to include in the migration equation for area i a variable for employment change in each region. Such a specification would be analogous to that used in our earlier paper (1984). A change in national employment can lead to no overall net migration within a closed system. This adding-up property must hold, and a specification such as that suggested here would guarantee that it would. Since we obviously have far too few temporal degrees of freedom to estimate a model that includes in the migration equation variables for employment change in each region, we have modified the approach by including a variable for employment change in all other regions, which is similar to the procedure used by William Milne (1981).

National productivity and real wage changes are to a greater or lesser extent shared by the nation's various regions. Interregional factor mobility also produces a tendency for the various regions to participate in national real wage growth. The variable for change in national average real earnings (ΔWUS) is thus included in the equation for change in local real earnings as a means of tying local wage changes to national wage changes.

A measure of change in aggregate local demand, namely ΔY, has been included in both the equation for employment change and that for wage change. The only usable measure we have is change in total annual earnings in Social Security-covered employment, which is expressed in constant dollars. Thus, the ΔE and ΔW equations contain variables for corresponding national conditions that are meant to proxy demand in nonlocal markets, as well as variables for local conditions that are meant to proxy demand in local markets. The income terms are obviously expected to take positive signs in both the employment and wage-change equations.

Finally, note that each behavioral equation contains a lagged endogenous variable. Since Social Security data are the only time-series data that are available for BEA Economic Areas for the 18-year period, a number of variables of potential relevance, such as area unemployment rates, are simply nonexistent. One strength of the temporal data is that they enable us to make use of the fact that these unobservables are already to some degree reflected in the past migration flows and employment and wage changes for a given region.

III. Empirical Results

The model described above is simultaneous in nature and includes lagged endogenous variables in each equation. Moreover, the error process tends to follow a first-order autoregressive scheme. Therefore, Ray Fair's (1970) method of estimation is used with the autocorrelation adjustment accomplished via the Hildreth-Lu scanning technique. We cannot, however, routinely appeal to the consistency and efficiency of the Fair estimators since these are asymptotic properties. Moreover, our small sample precludes a theoretically based choice of the autocorrelation estimator. Finally, the potential bias imparted by any zero restrictions necessitated by data limitations could conceivably lead to erroneous inferences. The potential for these problems to produce not only unreasonable individual parameter estimates, but also unreasonable dynamic behavior of the model given its partial-adjustment form, has led us to seek validation of the model for each BEA area through fully dynamic historical simulations and through several internal consistency checks on various model parameters. Validation procedures are discussed at a later point in the paper.

For the 57 areas that are the focus of this study, Table 1 reports the coefficients (α_1) on the employment variable in the net-migration equations and the coefficients (β_1) on the net-migration variable in the employment equations. In Table 1 the areas are listed from highest to lowest in terms of average employment size. The coefficients on the employment variable in the net-migration equation are almost all positive (56 of 57) and 54 are significantly different from

TABLE 1—MIGRANT-ATTRACTIVE POWER OF AN ADDITIONAL JOB AND NUMBER OF JOBS RESULTING FROM AN ADDITIONAL MIGRANT: FAIR METHOD ESTIMATES FOR 57 LARGE AREAS

Area[a]	α_1	β_1	Area[a]	α_1	β_1
New York	0.496 (> 0)	1.115 (=1)	Birmingham	0.457 (> 0)	1.421 (=1)
Los Angeles	0.679 (> 0)	0.506 (<1)	Tampa	0.475 (> 0)	1.374 (>1)
Chicago	0.214 (> 0)	0.213 (=1, =0)	Columbus	−0.012 (= 0)	0.143 (=1, =0)
Philadelphia	0.250 (> 0)	1.270 (=1)	Denver	0.315 (> 0)	1.214 (=1)
Boston	0.254 (> 0)	1.651 (=1)	Raleigh	0.382 (> 0)	0.936 (=1)
Detroit	0.280 (> 0)	0.237 (=1, =0)	Memphis	0.471 (> 0)	0.862 (=1)
San Francisco	0.473 (> 0)	1.143 (=1)	Syracuse	0.528 (> 0)	1.077 (=1)
Cleveland	0.362 (> 0)	1.777 (=1)	Nashville	0.803 (> 0)	0.892 (=1)
Pittsburgh	0.184 (> 0)	2.068 (=1)	Greensboro	0.513 (= 0)	0.794 (=1)
St. Louis	0.722 (> 0)	1.314 (>1)	Phoenix	0.531 (> 0)	1.362 (=1)
Hartford	0.512 (> 0)	0.616 (=1)	Louisville	0.439 (> 0)	1.606 (=1)
Minneapolis	0.328 (> 0)	2.281 (>1)	Huntington	0.454 (> 0)	1.189 (=1)
Dallas	0.453 (> 0)	1.314 (=1)	Lansing	0.513 (> 0)	1.331 (=1)
Washington, D.C.	0.516 (> 0)	1.520 (>1)	Grand Rapids	0.249 (= 0)	−0.402 (<1, =0)
Baltimore	0.531 (> 0)	0.397 (<1)	Richmond	0.670 (> 0)	1.426 (>1)
Miami	0.369 (> 0)	1.763 (=1)	Oklahoma City	0.446 (> 0)	1.581 (>1)
Atlanta	0.784 (> 0)	0.927 (=1)	Dayton	0.466 (> 0)	1.643 (>1)*
Houston	0.516 (> 0)	1.661 (>1)	Salt Lake City	0.479 (> 0)	1.985 (>1)
Milwaukee	0.166 (= 0)	1.079 (=1, =0)	Rochester	0.607 (> 0)	0.995 (=1)
Kansas City	0.421 (> 0)	0.927 (=1)	San Diego	0.553 (> 0)	1.688 (>1)
Seattle	0.449 (> 0)	1.396 (=1)	Toledo	0.519 (> 0)	1.236 (=1)
Buffalo	0.439 (> 0)	1.327 (=1)	Greenville, SC	0.786 (> 0)	0.809 (=1)
New Orleans	0.508 (> 0)	1.613 (>1)	Tulsa	0.420 (> 0)	1.831 (>1)
Cincinnati	0.462 (> 0)	1.717 (>1)	Norfolk	0.387 (> 0)	1.871 (=1)
Harrisburg	0.260 (> 0)	2.541 (>1)	Jacksonville	0.373 (> 0)	1.690 (> 0)
Indianapolis	0.601 (> 0)	1.002 (=1)	Sacramento	0.198 (> 0)	2.024 (=1)
Charlotte	0.624 (> 0)	0.741 (=1)	Roanoke	0.134 (> 0)*	1.447 (=1)*
Albany, NY	0.599 (> 0)	1.041 (=1)	San Antonio	0.556 (> 0)	1.650 (>1)*
Portland, OR	0.564 (> 0)	0.954 (=1)			

Notes: α_1 (= 0) indicates not significantly different from zero at the 0.975 level for a one-tail test; (> 0) indicates significantly greater than zero at the 0.975 level (*0.95) for a one-tail test. β_1 (= 0) indicates not significantly different from zero at the 0.975 level for a one-tail test; (=1) indicates not significantly different from 1.0 at the 0.95 level for a two-tail test; (>1) indicates significantly greater than 1.0 at the 0.95 level (*0.90) for a two-tail test.

[a]Only the first name is listed for areas with multiple cities in their title.

zero. The range of values is broad, running from a high of 0.803 (Nashville) to just less than zero. For each area, α_1 is significantly less than 1.0, which indicates that local residents absorb at least a fraction of any incremental employment. For the 57 areas shown in Table 1, the mean value of α_1 is 0.451.[9] The average absolute t-ratio is 5.837. This value of 0.451 is slightly less than the corresponding mean value of 0.487 for the 114 smallest areas, which are not discussed here. Note also that each of these values is somewhat lower than the 0.67 figure obtained by Muth.

The value of α_1 differs across the broad regions of the country. In terms of mean values of α_1, the regions rank as follows: 1) South, 0.503; 2) West, 0.471; 3) Northeast, 0.438; and 4) North Central, 0.365. Furthermore, the differences between the South and West relative to the Northeast and North Central regions are statistically significant. To demonstrate the significance of the difference, we estimated the following re-

[9]Values reported here and in Table 1 are for α_1, rather $(\alpha_1 + \alpha_2)$. Although α_2 is negative and significant for almost every area, it is small in absolute value relative to α_1. The absolute value of α_2 is far higher for New York and Los Angeles than for any other area, and the respective adjustments for these areas are only −0.052 and −0.037, respectively. Most of the other areas have values of α_2 closer to zero than −0.010.

gression, where D is a dummy variable that equals 1 if the area is in the Northeast or North Central regions and 0 if it is in the South or West:

$$\alpha_{1i} = \underset{(17.795)}{0.494} - \underset{(2.377)}{0.102D}, \tag{6}$$

$$R^2 = 0.093, \ F = 5.648.$$

Numbers in parentheses are t-ratios. Note that the coefficient on D is highly significant and indicates that 10 additional jobs in the average area in the Northeast and North Central regions will attract approximately one less migrant than the same number of extra jobs in the average southern or western area.[10]

The coefficients on the net-migration variables in the employment equations are almost all positive (56 of 57) and are generally significantly different from zero (52 of 57). A majority of the coefficients is consistent with Muth's earlier finding that migration results in a proportionate increase in employment. For 38 of the 57 areas (66.7 percent), the hypothesis that $\beta_1 = 1$ cannot be rejected. For 16 areas, however, the hypothesis that $\beta_1 > 1$ cannot be rejected. Thus, for almost 30 percent of the areas the in-migration of one employed person results in more than one additional job. For three areas (Los Angeles, Baltimore, and Grand Rapids) migrants appear to displace indigenous workers in the sense that employment rises by less than the migrant's contribution to it.

The mean value of β_1 for the 57 large areas is 1.259, with a corresponding mean absolute t-ratio of 4.630. The corresponding mean value for the 114 smallest regions is 1.113 (with $|t| = 3.871$). The mean values differ slightly across the four broad regions. These means are as follows: West, 1.364; South, 1.305; Northeast, 1.293; and North Central, 1.104. Although the mean values of β_1 favor the West and the South over the Northeast and North Central regions, the differences are not statistically significant. Because the results are symmetrical regarding net in- and out-migration, they also suggest that another employed net out-migrant from a large area results in an average decrease in employment of 1.259 jobs, or a decrease that exceeds his own job by 0.259. The net out-migration that has been observed from many large metropolitan areas over the last decade or so thus may have cumulative effects that are not insubstantial.

IV. Local Impacts of National Forces

As pointed out previously, the inclusion of a variable for growth of U.S. employment in the equation for area employment growth ties the temporal performance of employment in each metropolitan area to the temporal performance of the national economy. Each specific local area is assumed to contribute a sufficiently small fraction of national employment that the variable for growth of U.S. employment can be treated as exogenous in the equation for growth of local employment.

The 57 areas that are the focus of this study held an average annual share of 77.1 percent of national employment over the study period, and they are estimated to have gained 71.1 percent of a given number of incremental national jobs, which is the sum of the β_2s over the 57 large areas. As indicated in regression equation (7), the average share of incremental national employment that accrued to each metropolitan area is similar to the area's share of national employment:

$$\beta_{2i} = -\underset{(0.881)}{0.0006} + \underset{(29.282)}{0.9686 S_i}, \tag{7}$$

$$R^2 = 0.94, \ F = 857.5.$$

The dependent variable in the above regression is the estimated coefficient on ΔEUS in the equation for local employment change, and S_i is area i's average share of national employment over the study period. Values in parentheses are t-ratios.

Because the metropolitan areas located in the Northeast and North Central regions are

[10]When relationship (6) is estimated with the estimated value of $(\alpha_1 + \alpha_2)$ as the dependent variable, the coefficient on D (-0.104) remains highly significant ($t = -2.464$).

relatively large and thus hold relatively large shares of national employment, they account for significantly more direct national incremental employment than their counterparts in the South and West. This observation is confirmed by the following regression:

$$\beta_{2i} = \underset{(2.778)}{0.0076} + \underset{(2.704)}{0.0115}D, \quad R^2 = 0.12,\ F = 7.313. \tag{8}$$

Recall that D is a dummy variable that takes a value of 1 for areas located in the Northeast and North Central regions. Relationship (8) indicates that the average metropolitan area of the Northeast and North Central regions gains a share of direct incremental national employment that is about 1.15 percentage points greater than that gained by the average metropolitan area of the South and West.

What appears to be happening is that incremental national employment accrues to each region roughly in proportion to the region's share of national employment. Whereas the extra jobs in each region attract migrants, the extra jobs that accrue to other regions discourage migration from those regions to any given region, as indicated by the magnitude of α_2. Because internal migration is nationally a zero-sum proposition, whether a given region gains or loses migrants depends importantly upon the migrant attractive power of another job in the region, the extent to which migrants are discouraged from moving to the region due to the availability of additional employment opportunities elsewhere, and the number of local jobs that result from the in-migration of an additional employed person. Depending upon the magnitudes of these and other relationships inherent in the model, net migration for a given area could be positive or negative when national employment expands by a specific amount.

The reduced-form coefficient associated with ΔEUS in the equation for local net migration provides a measure of the direction and magnitude of an area's net migration that results from a given expansion of national employment. As shown by the following regression, southern and western regions tend to gain migrants when national employment increases, whereas northeastern and midwestern areas tend to lose them:

$$RF\alpha_{2i} = \underset{(3.182)}{0.0043} - \underset{(2.405)}{0.0050}D \quad R^2 = 0.095,\ F = 5.782. \tag{9}$$

In this regression $RF\alpha_2$ is the reduced-form coefficient associated with ΔEUS in the equation for local net migration, and D is again the regional dummy variable. The t-value shown under the coefficient associated with D indicates that regions in the South and West gain net migrants at the expense of regions in the Northeast and North Central parts of the country. Specifically, the average large metropolitan area located in the Northeast and North Central United States loses 5 employed net out-migrants for every 1,000 incremental national jobs.

Although this estimate may seem high, it is not completely out of line. During the sample period Social Security employment grew nationally by about 1,700,000 per year, which translates into 8,500 employed net out-migrants from the average metropolitan area of the Northeast and North Central regions. Since 26 such areas are included in these two broad regions, 221,000 employed net out-migrants are expected from all 26 taken as a whole. Over a 5-year period we could expect roughly 5 times this magnitude, or 1,105,000 employed out-migrants. Since the employment-to-population ratio is approximately 0.40, let us multiply 1,105,000 by 2.5, which yields 2,762,500 as a rough estimate of population migration out of the major metropolitan areas of the Northeast and North Central regions. Of course, some fraction of these net out-migrants may have moved to smaller areas of the Northeast and North Central regions, so net migration to the South and West from these 26 areas would be lower. Between 1970 and 1975, net migration between these two broad regions was estimated by the U.S. Bureau of the Census to be 2,537,000 persons.[11]

[11] See Greenwood (1981, p. 44).

Since internal migration causes a redistribution of employment gains, we should find that the pattern of reduced-form coefficients on the ΔEUS variable in the equations for local employment change indicates a lesser share of employment accruing to Northeast and North Central areas than suggested by the various values of β_2. Indeed, precisely this finding is observed. In the regression below, the dependent variable is the reduced-form coefficient on ΔEUS in the ΔE_t equation:

$$(10) \quad RF\beta_{2i} = \underset{(4.368)}{0.0166} + \underset{(0.052)}{0.0003}D, \qquad R^2 = 0.00, \ F = 0.003.$$

Note that the regional dummy variable, which was positive and statistically significant in the regression for β_{2i}, is now insignificant. Southern and western areas appear to capture shares of incremental national employment that are ultimately no smaller than those that accrue to northeastern and midwestern areas.[12] This observation suggests, of course, that the former regions gain a share of incremental employment somewhat greater than their existing share of national employment.

V. Model Validation

Although the data that underlie this analysis are unique, they present problems. We mentioned previously that the data relate only to Social Security-covered employment and that consequently to be classified as a migrant, an individual must be employed in a Social Security-covered job both before and after moving. To the extent that unemployed persons migrate and subsequently find employment in a Social Security-covered job, the reported values of α_1 are lower than they ought to be, whereas those of β_1 are higher than they ought to be.[13] Perhaps more importantly, however, only 18 annual observations are available on migration and a limited number of other variables of interest. Time-series data on many variables that are likely to importantly interact with migration simply do not exist for BEA Economic Areas. These types of data limitations result in two problems. First, the small number of available observations prohibits the estimation of parameters that embody well-known, large-sample properties. Consequently, uncertainty is present regarding the econometric properties of the estimated coefficients. Second, lack of appropriate data forces (zero) restrictions on the model that open it to question regarding specification.

To validate the internal consistency of the model, we have taken a number of additional steps. As mentioned earlier, certain restrictions must logically hold across the entire set of 171 areas, but these restrictions could not be explicitly built into the model due primarily to lack of sufficient temporal observations. The most obvious restrictions of this type are

$$(11) \quad \sum_{i=1}^{171} NM_{it} = 0 \quad \text{for all } t,$$

$$(12) \quad \sum_{i=1}^{171} RF\alpha_{2i} = 0,$$

$$(13) \quad \sum_{i=1}^{171} \beta_{2i} = 1,$$

$$(14) \quad \sum_{i=1}^{171} RF\beta_{2i} = 1.$$

[12] The relationship between $RF\beta_{2i}$ and S_i (the area's average share of national employment) is also somewhat weakened compared to that between β_{2i} and S_i, as indicated by the following regression:

$$RF\beta_{2i} = \underset{(1.747)}{0.0048} + \underset{(6.922)}{0.8847}S_i, \qquad R^2 = 0.47, \ F = 47.911.$$

[13] Data from the March 1973 *Current Population Survey* make possible at least a crude adjustment for unemployed migrants. In the *CPS*, however, migration is defined in terms of counties. Samuel Saben (1964) reports that of 2,267,000 migrants who were employed in March 1962, 2,031,000 were employed in March 1963. Of 235,000 migrants who were unemployed in March 1962, 169,000 were employed in March 1963. Thus, of those migrants who were employed at the end of the period, about 7.68 percent were unemployed at the beginning of the period. If we use 1.0768 as a crude adjustment factor, the estimate of the mean value of α_1 rises from 0.451 to 0.486, while that of β_1 falls from 1.259 to 1.169.

In other words, internal migration is analogous to a zero-sum game, as indicated by (11). Consequently, the reduced-form coefficients associated with ΔEUS in the various equations for area net migration must sum to zero, as suggested by (12). Relationship (13) indicates that any increment in national employment must be directly shared exhaustively among the 171 regions, and relationship (14) suggests that when all of the interactions inherent in the model are taken into account, national employment cannot rise by more than the given increment during the period, which is one year. The various values of $RF\beta_{2i}$ show the shares of national incremental employment that accrue to each region after the interactions have been taken into account. If, in practice, these various restrictions fail to hold approximately, then the internal consistency of the estimated model would be open to serious question, and our ability to draw national inferences from the aggregate of individually estimated models would also be doubtful. Let us thus consider how closely each restriction comes to holding.

Net migration must sum to zero nationally, but in fact $\Sigma RF\alpha_{2i} = 0.117$, which indicates that an average of 0.00068 net inmigrants must be deducted from each area's reduced-form coefficient to center national net migration on zero. If national employment were to expand by one million jobs, 680 migrants would thus have to be substracted from each area to guarantee logical consistency—that is, to guarantee that $\Sigma NM_i = 0$. This number is not particularly large compared to the overall migration impacts on the large areas. For example, rather than losing 3,363 net out-migrants when national employment expands by one million jobs, the average area that lost migrants under such circumstances would lose 4,043. Rather than gaining 6,852 employed inmigrants under such circumstances, the average area that experienced increased migration due to national employment expansion would gain 6,172. Of course, other means of distributing this error, such as in proportion to area size, are available.

When β_{2i} is summed over the entire set of 171 areas, it fails to satisfy (13) by 4 percent; that is, $\Sigma\beta_{2i} = 0.960$, which is not statistically different from 1.0. Moreover, $\Sigma RF\beta_{2i} = 1.068$, which suggests that the sum of the reduced-form coefficients is about 7 percent higher than it should be. Although the various restrictions are not satisfied exactly, they are quite close to what they ought to be.

As a means of further validating the model results, we performed a number of simulation experiments. The procedure was as follows. First, for selected areas, the initial values of the endogenous and exogenous variables were assigned to the estimated models, along with the historic series on the exogenous variables (i.e., ΔEUS, ΔWUS, and RW_{t-1}), and historical simulations were generated. Although we experimented with the use of a number of different algorithms for the solution of nonlinear systems and also with convergence criteria, the models never converged to a solution. The nonlinear identity was clearly troublesome. Thus, we replaced this identity with a linear behavioral equation (i.e., $\Delta Y = f(\Delta\hat{E}, \Delta\hat{W})$). (Note that this substitution in no way affects the values of the estimated parameters in other equations.) The procedure allowed us to both simulate the models and derive reduced-form coefficients for each exogenous variable, which was accomplished for all 171 areas, as noted above.

Little information is available on simulations of models such as that presented here, primarily because the lack of time-series data on migration has prevented the development of temporal models that incorporate this variable. Thus, we have little to which to relate our simulation results. David Reaume's (1983) recent study is an exception. He reports that in his simulations of 4 state models that include indirect measures of net migration, the fitted models are explosive in every case, due mainly to the estimated migration parameters. Our results are much more encouraging. For example, the root mean square percentage error between the actual and simulated employment levels fails to exceed 10 percent for 27 areas (47 percent of the cases) and does not exceed 50 percent for 37 areas (65 percent of the cases). For less than one-third of the areas the root mean square percentage error exceeds 100 percent. For the wage series the results are slightly better. The root mean square per-

centage error is less than 10 percent for over 61 percent of the areas and is over 100 percent in less than 25 percent of them. Given the typical stability of the estimated models, we feel that the conclusions presented above are not overreaching. However, in light of the rather large root mean square percentage errors for certain areas, they should probably be regarded as provisional.

VI. Summary and Conclusions

The major objectives of this paper were to measure the migrant-attractive power of another job and the number of local jobs attributable to another migrant. In accomplishing these objectives we have employed unique time-series data on migration and employment growth for 171 U.S. regions. For the country's 57 largest metropolitan areas, the results indicate that an average of 0.451 employed net migrants are directly attracted by one additional job. Areas located in the South and West, however, attract about one more employed in-migrant per 10 extra jobs than those located in the Northeast and North Central United States. Moreover, for two-thirds of the nation's major metropolitan areas, one more employed in-migrant results in one additional job, which is consistent with Muth's earlier finding. For about 30 percent of the major urban centers, however, the in-migration of one additional employed migrant results in more than one additional job. An extra net in-migrant to the average area is directly responsible for about 1.259 more jobs in the area, one job being his own and 0.259 accruing to others in the locality of in-migration. No statistically significant interregional difference exists in this relationship.[14] Since the latter relationships are symmetrical regarding net in- and net out-migration, they suggest that out-migration from large metropolitan areas, such as has been observed over the last decade, has substantial cumulative effects.

To an important extent the local employment impacts of national employment changes are greater where the migrant-attractive power of another job is greater and/or where the number of jobs resulting from another migrant is greater. What appears to be happening is that most areas gain a direct share of incremental national employment that is roughly proportional to the area's share of national employment. Because the large metropolitan areas of the Northeast and North Central United States hold relatively large shares of national employment, relatively much of a given increment of national employment initially accrues to them. However, because the migrant-attractive power of another job is greater in southern and western areas, migrants are drawn to these areas and away from those of the Northeast and North Central regions. Migrant induced employment thus tends to accrue to the regions of net in-migration at the expense of the regions of net out-migration. When these employment-migration interactions are taken into account, national incremental employment accrues more equally to the South and the West compared

[14] This statement refers to the lack of significance of a regional dummy variable in a regression whose dependent variable is our estimate of migration's direct impact on local employment. Another issue of considerable interest involves an analysis of the spatial (or cross-sectional) variation in the temporally estimated coefficients. Although such an analysis is beyond the scope of this paper, let us briefly consider some reasons for the variation. Preliminary analysis has involved the estimation of generalized least squares relationships, where the weights are given by the standard errors associated with the original estimates of α_1 and β_1. These results suggest that the number of jobs that directly result from another migrant (β_1) is greater for areas with higher average rates of employment growth over the study period, presumably because such rates proxy the vintage of area's physical and human capital stocks, and consequently labor productivity. This relationship holds within both broad regions. Within the South and West, but not the Northeast and North Central regions, β_1 falls as area size (measured as average employment level over the study period) rises. Moreover, within the South and West the migrant attractive power of an incremental job (α_1) varies directly with area size and inversely with size squared. Furthermore, α_1 is lower where the average unemployment rate (i.e., the average of 1960 and 1970 census unemployment rates) is higher. These relationships (for α_1) fail to hold within the Northeast and North Central region.

to the Northeast and North Central regions, which means that the former regions gain a disproportionate relative share of such employment. This finding appears to be consistent with an observation made by Muth some years ago, but largely overlooked since: "The total employment multiplier, rather, depends upon the extent to which an increase in employment in excess of natural increase in the labor force induces in-migration" (1968, p. 313).

In other respects our results are also quite close to Muth's earlier findings, and where they differ, they do so in the expected directions. For example, our estimates of the migrant attractive power of an incremental job are similar to Muth's, but generally lower. This difference is in the expected direction since the present results are derived from annual data, whereas Muth's are derived from data measured over a 10-year period. Moreover, our findings in general suggest that for large metropolitan areas labor demand is highly elastic and labor supply is highly inelastic. Muth's findings imply the same general configurations of labor demand and supply relationships.

The findings reported here appear to have important policy implications. One of the most frequently suggested proposals for dealing with the problems of the old industrial areas of the Northeast and North Central regions is strong fiscal and monetary policies that would presumably have relatively sizeable effects in the decaying areas of the country. The results of this study suggest that such policies would have positive impacts in the Northeast and North Central regions, but the flow of population out of those areas may not be discouraged. On the contrary, because the incremental jobs resulting from expansionary national policies are more attractive to migrants if they occur in the South and the West, further migration to these latter regions may even be encouraged. By implication, our results emphasize the importance of targeting relief to eastern and northern areas, if the maintenance of these areas at former levels of relative prominence is deemed in the national interest.

REFERENCES

Borts, George H. and Stein, Jerome L., *Economic Growth in a Free Market*, New York: Columbia University Press, 1964.

Fair, Ray C., "The Estimation of Simultaneous Equation Models with Lagged Endogenous Variables and First Order Serially Correlated Errors," *Econometrica*, May 1970, *38*, 507–16.

Goldstein, Gerald S. and Moses, Leon N., "A Survey of Urban Economics," *Journal of Economic Literature*, June 1973, *11*, 471–515.

Greenwood, Michael J., *Migration and Economic Growth in the United States: National, Regional, and Metropolitan Perspectives*, New York: Academic Press, 1981.

______ **and Hunt, Gary L.,** "Econometrically Accounting for Identities and Restrictions in Models of Interregional Migration," *Regional Science and Urban Economics*, February 1984, *14*, 113–28.

Mazek, Warren E., and Chang, John, "The Chicken or Egg Fowl-Up in Migration: Comment," *Southern Economic Journal*, July 1972, *39*, 133–39.

Milne, William J., "Migration in an Interregional Macroeconometric Model of the United States: Will Net Out-Migration from the Northeast Continue?," *International Regional Science Review*, Fall 1981, *6*, 71–83.

Muth, Richard F., "Differential Growth among Large U.S. Cities," in J. P. Quick and A. M. Zarley, eds., *Papers in Quantitative Economics*, Lawrence: University Press of Kansas, 1968, 311–55.

______, "Migration: Chicken or Egg?," *Southern Economic Journal*, January 1971, *37*, 295–306.

Reaume, David M., "Migration and the Dynamic Stability of Regional Econometric Models," *Economic Inquiry*, April 1983, *21*, 281–93.

Saben, Samuel, "Geographic Mobility and Employment Status, March 1962–March 1963," *Monthly Labor Review*, August 1964, *87*, 873–81.

Endogenous Tariff Formation

By WOLFGANG MAYER*

Interest in the process of tariff formation has grown considerably during the last decade. Robert Baldwin (1976, 1982), William Brock and Stephen Magee (1978, 1980), as well as Ronald Findlay and Stanislaw Wellisz (1982) have suggested alternative models for studying the economic and political forces that interact in determining a country's actual tariff structure.[1] The underlying premise of these studies is that political decisions on tariff rates are reflections of the selfish economic interests of voters, lobbying groups, politicians, or other decision makers in trade policy matters.[2,3]

Brock and Magee, as well as Findlay and Wellisz, treat tariff formation as a noncooperative game among competing economic interest groups.[4] In Brock and Magee's papers, the tariff positions of opposing political parties are explained, where each party maximizes its chances of being elected, and the probability of reelection hinges on contributions from tariff-sensitive lobbying groups. In Findlay and Wellisz, the economic interests of land and capital owners are opposed, with labor standing on the sidelines. Baldwin, on the other hand, postulates majority voting by owners of productive factors. Hence, Baldwin introduces a collective decision rule to tariff determination which plays a major role in the public choice literature.[5] Specifically, he argues that accepted optimal trade policy prescriptions, such as free trade for a small country, would be voted in by factor owners if voters' tastes were homothetic and there were no restrictions on income redistributions, no voting and information costs, and no possibilities for logrolling. However, when these ideal assumptions are gradually removed, Baldwin (1976, p. 71) suggests that actual tariff rates are decisively affected by economic groups which are large in size and whose potential gains or losses are substantial.

Baldwin's discussion is rich in detail and offers many valuable insights. Most important, it breaks new ground in suggesting determinants of actual tariffs, such as the distribution of factors of production, the existence of voting costs, and the specific nature of the underlying economic and political system. The purpose of this paper is to study some of these alternative determinants of tariff rates in a rigorously formulated general equilibrium model. In particular, the paper attempts to evaluate the dependence of actual tariff rates on factor-ownership distribution, voter eligibility and participation rules, and the degrees of factor mobility and industry diversification in the economy.

The standard assumption on factor ownership, which also is adopted by Baldwin (1982, p. 268), states that each person owns one factor of production only. This implies that all owners of a given factor form a

*Department of Economics, University of Cincinnati, Cincinnati, OH 45221. Comments by John Pomery, Raymond Riezman, and Ronald Jones are greatly appreciated.

[1] In addition to the literature on endogenous tariff formation, there are a number of authors (for example, Jagdish Bhagwati, 1982; Arye Hillman, 1982; and Elias Dinopoulos, 1983), who discuss endogenous tariff adjustments in response to exogenous changes in the international terms of trade. How the initial tariff rate was selected is, however, not part of their decision problem. Also there is a growing empirical literature on changing tariff structures of various countries, such as the works by Jonathan Pincus (1975), Richard Caves (1976), G. K. Helleiner (1977), as well as Howard Marvel and Edward Ray (1983).

[2] The relationship between public policymaking and economic interests is widely discussed in the public choice literature, such as in Anthony Downs (1957), James Buchanan and Gordon Tullock (1962), and Dennis Mueller (1979).

[3] A quite different approach at explaining tariff structures has been used by Edward Ray (1974).

[4] For extensions of these modes, see Leslie Young and Magee (1983), Wellisz-Findlay (1983), and Wellisz and John Wilson (1983).

[5] For a thorough discussion of majority voting rules, see Duncan Black (1948a,b) and Mueller.

well-defined interest group whose membership is independent of the prevailing tariff rate. Accordingly, voter allegiances are exogenously imposed rather than endogenously determined. Baldwin's correct observation, that majority voting could eliminate all trade of a capital-abundant country if there were more workers than capitalists, suggests that this standard assumption on factor ownership has limited appeal in explaining actual tariff structures under majority voting. This paper changes the standard assumption by allowing a person to own more than one factor of production. Also, factor-ownership shares may differ among people. As a result, the number of factor owners voting for or against a given tariff change will no longer be fixed, but will depend on the prevailing tariff rate. Under shifting voter allegiances, a tariff equilibrium is attained when no majority can be found to push for either an increase or a reduction in the existing tariff rate. Whether such an equilibrium tariff is positive, zero, or negative crucially hinges on the underlying factor-ownership distribution. I demonstrate that each factor owner has an optimal tariff rate whose value is uniquely related to the individual's factor ownership. In the special case of majority voting with no voting costs, it is the median factor owner's optimal tariff rate that will be chosen to become the actual tariff rate.

Given a country's distribution of factor ownership, tariff equilibria are affected by voter eligibility rules and participation costs. Some factor owners may not be eligible to participate in the voting process. Others are eligible but have no incentive to participate as voting costs are high relative to benefits from enacting or preventing a proposed tariff change.[6] Over the last century, most countries have undergone drastic liberalizations with respect to voter eligibility rules. But even today there are many factor owners, such as young people or foreign workers, who are not permitted to vote. This paper also relates actual tariff structures to voter eligibility rules and, more importantly, to the existence of significant voting costs on the part of factor owners.

Tariff formation is discussed under two alternative assumptions about the underlying production structure. First, I employ the common two-by-two Heckscher-Ohlin model, later a multisectoral factor-specific model is introduced. The advantage of the former is that it is sufficiently simple to illustrate the tariff formation process and it offers a plausible explanation for long-run tariff adjustments as factor-ownership distribution, voting costs, or voter eligibility rules change. Its main shortcoming is its failure to explain the frequently observed phenomenon that a relatively small industry, that does not have support from the majority of eligible voters, succeeds in gaining tariff protection. The factor-specific multisector model is more appropriate for studying such industry-specific efforts to raise a given tariff. In such a model, higher tariffs on a given import good lead to significant welfare gains for the average specific-factor owner in the protected industry but to rather small welfare losses for average specific-factor owners in all other industries. The small number of big potential gainers, therefore, has much greater incentives to participate in the political process than the large number of small potential losers, whenever there exist significant voting costs. Hence, I am proving rigorously what Baldwin (1976, p. 71) hypothesizes and Anthony Downs suggested, namely that small important producer groups can be quite successful in securing import protection.

In addition to studying the determinants of actual tariff rates, I will ask whether there exist factor-ownership distributions under which the majority of voters would support a free-trade policy for a small country. Stated differently, I am in search of those factor-ownership distributions that, given the assumed production structure, result in Pareto-efficient resource allocations.

Section I states general assumptions about consumer preferences, factor distributions, and tariff revenue redistributions in a Heck-

[6] In the tariff formation literature, voting costs have been emphasized by Baldwin. More generally, they play an important role in discussions of the voter's calculus; for example, Downs, William Riker and Peter Ordeshook (1968), or Mueller (pp. 121–24). Sam Peltzman (1976) incorporates voting costs into his general theory of regulation.

scher-Ohlin model, and describes what tariff rate is optimal for a given person with predetermined factor ownership. Section II deals with the impact of tariff changes on the individual's economic interests, as measured by real income changes, and demonstrates how the actual tariff rate is chosen under majority voting. The relationship between the chosen tariff rate and changing voting costs is emphasized in this context. Section III introduces a multisector, factor-specific production model, as discussed in Ronald Jones (1975) and expanded in Roy Ruffin and Jones (1977). My aim is to demonstrate that, under rather neutral assumptions, there is a very strong presumption that a small industry will succeed in gaining import protection even though all other industries are hurt.

I. Optimal Tariff Rates in a Heckscher-Ohlin Model

A. *Assumptions*

Consider a small open economy in which capital and labor are employed to produce two commodities, X_1 and X_2. Each inhabitant of this country possesses one unit of labor and no or some positive amount of capital; that is, $L^i=1$ and $K^i \geqq 0$, where $L^i(K^i)$ denotes labor (capital) owned by individual i, $i=1,\dots,I$. Both factors are perfectly mobile between the two industries, all markets are competitive, and the firms' production functions are homogeneous of degree one.

Preferences of the country's inhabitants are assumed to be homothetic and identical,[7] such that a redistribution of income will affect neither the country's aggregate demand nor imports, as long as total income remains unchanged. Preferences of individual i are described by an indirect utility function,

$$(1) \qquad U^i = U^i(p, y^i), \qquad i=1,\dots,I,$$

[7]Furthermore, the ith individual's direct utility function is assumed to be homogeneous of degree one in the two commodities.

where U^i denotes maximum utility attainable by individual i, given the price of the first in terms of the second commodity, p, and income of individual i, y^i, measured in terms of the second commodity. There are three potential sources of income: ownership of labor, ownership of capital, and redistributions from tariff revenues. Total income of individual i is

$$(2) \qquad y^i = w + rK^i + T^i,$$

where w and r denote the returns on labor and capital, respectively, T^i represents tariff revenues received by individual i, and $L^i=1$. Concerning the redistribution of tariff revenues, it is assumed that there are predetermined rules that are not voted on in conjunction with the tariff issue, and that are neutral with respect to the overall distribution of income. Neutrality means that the ith individual's share of total tariff revenues is identical to the ith person's income share from factor ownership, ϕ^i; that is,

$$(3) \qquad T^i = \phi^i T,$$

where T represents total tariff revenues and

$$(4) \qquad \phi^i = (w + rK^i)/(wL + rK).$$

In equation (4), L is the total labor supply or number of inhabitants and $K = \sum_i^I K^i$. Also, it should be noted that $\sum_i^I \phi^i = 1$. Substituting (3) and (4) in (2) yields

$$(2') \qquad y^i = \phi^i(wL + rK + T) = \phi^i Y,$$

where Y denotes the economy's total income in terms of the second commodity.

Finally, it is assumed that, for all relevant tariff rates, incomplete specialization in production prevails and the first commodity is imported. Total tariff revenues are expressed as

$$(5) \qquad T = t\pi M,$$

where t is the tariff rate, π the world price, and M the imported quantity of the first commodity.

B. *Optimal Tariff Rates for Individuals*

When factor ownership differs among a country's people, their welfare is not affected uniformly by a tariff increase, as can be seen by substituting (2′) in (1):

$$(1') \qquad U^i = U^i(p, \phi^i Y), \qquad i = 1, \dots, I.$$

The ith person's maximum attainable utility depends on its own income share, in addition to domestic prices and aggregate income as faced by all individuals. If a tariff rate is changed, all people are confronted with identical price and aggregate income changes, but in general with different effects on their income share. Consequently, a given tariff change benefits some and hurts others, whereby the magnitude of welfare changes may vary greatly among gainers and losers. Inequality in the factor-ownership distribution is responsible for this differentiation in welfare effects.

To see the impact of a tariff increase on a given individual's welfare, let us substitute $p = \pi(1+t)$ in (1′) and differentiate (1′) with respect to t:

$$(6) \qquad (\partial U^i/\partial t) = (\partial U^i/\partial y^i) \times \left[-\phi^i \pi D_1 + \phi^i(\partial Y/\partial t) + Y(\partial \phi^i/\partial t)\right],$$

where D_1 denotes aggregate demand for the first commodity. In deriving (6), use is made of Roy's Identity and of the property of homothetic utility functions that the ith person's demand for a commodity equals the product of its income share and aggregate demand for the same commodity.

The economy's aggregate income is given by

$$(7) \qquad Y = wL + rK + T = pX_1 + X_2 + t\pi M,$$

where X_j indicates industry output of commodity j, $j = 1, 2$. Differentiating (7) with respect to t yields

$$(8) \qquad (\partial Y/\partial t) = \pi D_1 + t\pi(\partial M/\partial t),$$

where $\partial M/\partial t < 0$. After substitution of (8), equation (6) can be rewritten as

$$(6') \qquad (\partial U^i/\partial t) = (\partial U^i/\partial y^i) \times \left[\phi^i t\pi(\partial M/\partial t) + Y(\partial \phi^i/\partial t)\right].$$

There are two channels through which individual welfare is altered by a tariff increase: changes in the tariff-weighted value of imports, $t\pi(\partial M/\partial t)$, and changes in the individual's income share, $(\partial \phi^i/\partial t)$. While imports always decline as import tariffs rise,[8] the direction of change of a person's income share depends on individual factor endowments relative to the nation as a whole, on the one hand, and on the relative factor intensity of the import industry, on the other hand. More will be said on this point shortly.

A tariff rate is optimal for individual i if $(\partial U^i/\partial t) = 0$ in (6′) and U^i is strictly concave in t.[9] Assuming the concavity property is satisfied, the optimal tariff rate for individual i, $\tilde{t}^i$, is given by

$$(9) \qquad \tilde{t}^i = -\left[Y/(\pi \partial M/\partial t)\right]\left[(\partial \phi^i/\partial t)/\phi^i\right].$$

Since $-(\partial M/\partial t) > 0$, a person's optimal tariff rate is positive, zero, or negative as a higher tariff raises, keeps constant, or lowers the person's income share.

The relationship between a tariff and a given person's income share, in turn, depends on that person's endowments, as well as on the production structure through which factor returns and commodity prices are linked. For example, in the present section, the economy's production side is described by a Heckscher-Ohlin model. Differentiating the income-share expression of (4) with respect to t, we obtain

$$(10) \qquad (\partial \phi^i/\partial t) = \left\{[wL]/\left[(wL + rK)^2(1+t)\right]\right\} \times \left\{r(k - k^i)(\hat{w} - \hat{r})/\hat{p}\right\},$$

[8] For a precise expression, see Jones (1969).

[9] Implicitly it is assumed that the import tariff is the only instrument to alter a person's welfare. Hence, the question of what the first-best instrument would be to redistribute income is not addressed in this paper.

where $k = K/L$ and $k^i = K^i/L^i = K^i$. The "hat" indicates percentage changes and $(\hat{w} - \hat{r})/\hat{p}$ is positive (negative) as the first or import commodity is relatively labor- (capital) intensive in production. Equation (10) reveals that a tariff increase results in a higher (lower) income share for the ith individual if the individual, compared to the nation as a whole, is relatively well (poorly) endowed with the import good's intensively used factor of production; that is, $\partial\phi^i/\partial t > 0$ if either $k^i < k$ and $k_1 < k_2$ or $k^i > k$ and $k_1 > k_2$, where k_j is the capital-labor ratio in industry j and good one is imported.

In light of this relationship between income shares and tariffs, the following conclusions can now be drawn about individually optimal tariff rates, as stated in (9):

1) The optimal import tariff is positive (negative) for people who are relatively well (poorly) endowed with the import good's intensively used factor.

2) The greater the difference between individual and national endowment ratios, the greater the deviation of individually optimal tariff or subsidy rates from free-trade policy.[10]

3) The optimal tariff rate is zero for each person whose personal capital-labor ownership ratio equals the national capital-labor endowment ratio.

II. Tariff Determination Under Majority Voting

Actual tariffs are assumed to be the result of majority voting, where votes reflect the economic interests of those that are eligible and willing to participate in the tariff formation process. This section starts with a description of the relationship between a person's economic interests, on one side, and actual tariff rates, as well as personal factor endowments, on the other side. It is followed by an analysis of tariff equilibria under alternative majority voting rules.

[10]If each person owned only labor or only capital, as usually assumed, there would be just one optimal trade policy for the "worker" and another one for the "capitalist," irrespective of individual persons' factor-ownership shares.

A. *Real Income Effects of Tariff Changes*

Changes in real income indicate to what extent a person's economic interests are affected. Measured in units of the second commodity, the real income change of individual i brought about by a tariff increase can be expressed as

$$(11)\quad B^i(k^i, t) = (\partial U^i/\partial t)/(\partial U^i/\partial y^i),$$

where $\partial B^i/\partial k^i > 0$ (< 0) if the imported good is relatively capital (labor) intensive and $\partial B^i/\partial t < 0$ always.

In order to show that $\partial B^i/\partial k^i > 0$ for a capital-intensive import good, let us first determine the value of the individual endowment ratio, $\tilde{k}^j$, at which the prevailing tariff rate t would be optimal. Using (9) and (10), the value of $\tilde{k}^j$ must be such that

$$(9')\quad t = -\left[\frac{Y}{\pi(\partial M/\partial t)}\right] \times \left[\frac{wL(\hat{w} - \hat{r})}{(wL + rK)\hat{p}(1+t)}\right]\left[\frac{r(k - \tilde{k}^j)}{(w + r\tilde{k}^j)}\right].$$

Next, we return to the definition of B^i and employ (6′) to obtain

$$(11')\quad B^i(k^i, t) = \phi^i\left[t\pi(\partial M/\partial t) + Y(\partial\phi^i/\partial t)/\phi^i\right].$$

Substitution of (9′) for t and of (10) for $(\partial\phi^i/\partial t)$, as well as some manipulations, yield

$$(11'')\quad B^i = -\left[\frac{(\hat{w} - \hat{r})}{\hat{p}(1+t)}\right] \times \left[\frac{Ywr}{(w + r\tilde{k}^j)(wL + rK)}\right][k^i - \tilde{k}^j],$$

where $-[(\hat{w} - \hat{r})/\hat{p}] > 0$ for a capital-intensive import good. Equation (11″) expresses that a given tariff increase raises (lowers) the ith person's real income if the ith person's capital-labor endowment ratio, k^i, exceeds

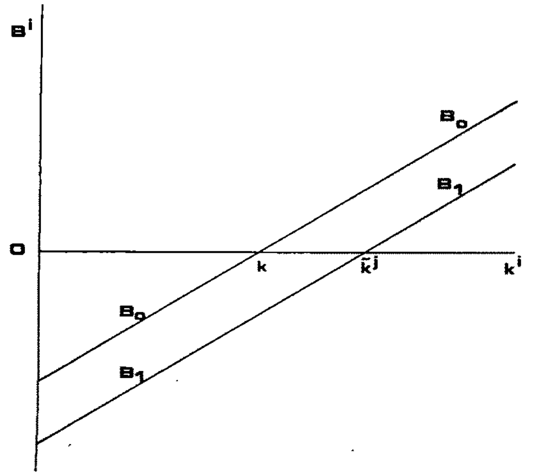


FIGURE 1

(falls short of) the capital-labor endowment ratio, $\tilde{k}^j$, of the person for whom the prevailing tariff rate is optimal. Furthermore, $(\partial B^i/\partial k^i)$ must be positive since, at a given tariff rate, no term inside the first two brackets of (11″) will change as k^i varies.

The negative sign of $(\partial B^i/\partial t)$, on the other hand, immediately follows from the utility function's properties of strict concavity in tariff rates and homogeneity of degree one in commodities.

The relationship between B^i, on the one hand, and k^i and t, on the other hand, is depicted in Figure 1 for an economy whose import industry is relatively capital intensive. The B_0B_0 locus is drawn under the assumption that there is free trade. Since $t=0$ when $\tilde{k}^j = k$, it follows from (11′) that $B^i \gtreqless 0$ as $k^i \gtreqless k$. The B_0B_0 locus is upward sloping as real income gains from tariff increases rise with the magnitude of a person's capital ownership. The B_1B_1 locus, on the other hand, is drawn under the assumption that the actual tariff rate is positive, that is, $t > 0$. This tariff rate would be optimal for a person with endowment ratio $\tilde{k}^j$. Only those individuals gain from further tariff increases for whom $k^i > \tilde{k}^j$, while all others lose. Consequently, if the level of tariffs is raised, the number of people who gain from further tariff increases becomes smaller, while the number of people who lose becomes larger.

B. *Majority Voting without Restrictions*

A tariff rate is called an equilibrium tariff rate when no majority of voters can be formed to alter that rate. Duncan Black (1948a) demonstrated that, in the absence of voting costs, such an equilibrium exists if voters possess "single-peaked" preferences and that the adopted policy is determined by the median voter's peak preference. Black's equilibrium condition is indeed satisfied in my tariff choice model, as each factor owner has a unique optimal tariff rate. Accordingly, the median voter's optimal tariff rate, denoted by $\tilde{t}^m$, becomes the equilibrium tariff rate under majority voting.

Who the median voter is depends on voter eligibility rules and the distribution of factor ownership. Concerning eligibility, assume at this point that there are no restrictions; each factor owner has one vote and participates in voting. Concerning factor ownership, it is assumed that capital is unevenly distributed, ranging from no capital ownership, $k^i = 0$, to maximum capital ownership, $k^i = k^{\max}$. It is convenient to describe capital-labor ownership ratios by $k(e)$, where e is an index such that $0 \leq e \leq 1$. All people with no capital ownership are indexed with $e = 0$, and e rises as the capital-labor ownership ratio rises; that is, $k(0) = 0$, $\partial k/\partial e > 0$, and $k(1) = k^{\max}$. The proportion of factor owners with the same factor ownership is defined as $f[k(e)] \geq 0$, where $\int_0^1 f[k(e)]\,de = 1$. As

$$(12) \qquad \int_0^1 k(e) f[k(e)]\, de = K/L = k,$$

the economy's capital-labor endowment ratio is the mean of the factor-ownership distribution. Equations (9) and (10) revealed that a given person's optimal tariff rate hinges on the relation between its own and the whole economy's capital-labor ratio. Applied to the median voter whose preferences are crucial for the actual choice of tariff rates, one can evaluate the *sign* of the economy's equilibrium tariff rate, t^*, by comparing the *median* with the *mean* of the factor-ownership distribution. Provided the distribution is unimodal,

$$(13) \qquad t^* = \tilde{t}^m \gtreqless 0,$$

as the distribution is skewed to the left, symmetric, or skewed to the right. A tariff (subsidy) on capital-intensive imports will be voted in if the majority of people has capital-labor ownership ratios that exceed (fall short of) the capital-labor endowment ratio of the country. And a small country will choose free trade as its policy in the special case of a symmetric distribution.[11,12]

For most nonsocialist countries there is strong evidence that capital-labor ownership distributions are not symmetric, but skewed to the right. Accordingly, one would expect a built-in tendency towards protection of labor's interests, through subsidies on capital-intensive imports or tariffs on labor-intensive imports. However, this intrinsic bias in favor of labor may not translate into prolabor tariff legislation if it is negated by voter eligibility and participation restrictions, as discussed next.

C. *Majority Voting under Restrictions*

Downs, Buchanan and Tullock, Baldwin, Peltzman, and many others have emphasized that there are considerable costs to participants in the voting process.[13] Voter participation costs imply that some eligible voters may choose not to exercise their voting rights. If participation costs exceed the gains from a proposed tariff change, an individual is best off not to vote at all. But even if gains from proposed tariff actions exceed voting costs, voter participation is not certain. Following Peltzman, this paper assumes that the probability of a person voting (or the fraction of people with identical factor ownership which will participate) is directly related to the magnitude of "net" gains.[14]

A person with capital-labor ownership index e will enjoy a net gain from a tariff increase if, for a given t,

$$B[k(e),t]-c(e)>0, \tag{14}$$

where $c(e)$ denotes the real income loss from voting, or simply voter participation costs, for people with factor-ownership ratio $k(e)$. The fraction of people with $k(e)$ that will vote for a tariff increase, $\rho(e)$, is

$$\rho(e)=\rho\{B[k(e),t]-c(e)]\}, \tag{15}$$

where $0\leq\rho<1$ and $\rho'\geqq 0$. Since the total number of people with factor ownership $k(e)$ equals $f[k(e)]L$, the total number of votes for a *tariff increase*, E, is

$$E=L\int_{\tilde{e}}^{1}\rho(e)f[k(e)]\,de, \tag{16}$$

where $\tilde{e}$ is the factor-ownership index for persons with zero net gains from the tariff rise; that is, for those with $B[k(\tilde{e}),t]=c(\tilde{e})$.

Similarly, a person will enjoy a net gain from a tariff *decrease*, if

$$D[k(e),t]-c(e) = -B[k(e),t]-c(e)>0, \tag{17}$$

where D is the real income gain from a tariff decrease. The fraction of people with factor-ownership ratio $k(e)$ that vote for the tariff

[11] Hence, if the capital-labor ownership distribution is symmetric and there are no voting restrictions, majority voting results in a Pareto-efficient allocation of resources.

[12] Under the assumptions of this model—one issue, all individuals vote, symmetric, unimodal factor distribution—free trade would also be chosen in a two-party representative democracy.

[13] The terms "voting" and "voter participation costs" have a broader meaning than marking a ballot and incurring costs in going to the voting booth. Voting can be interpreted as any form of communicating a person's response to a given policy proposal whether through the ballot, letters to political representatives, participation in political meetings, or public demonstrations. Since a politician in a two-party representative democracy will also choose the median-voter's optimal tariff, informing politicians of one's attitudes is a form of voting. Voting costs include all expenses incurred in transmitting this information. In parts of the public choice literature, voting costs also include information-gathering costs. Strictly speaking, this is legitimate only if B^i is independent of such information-gathering costs.

[14] Peltzman's formulation of voting probabilities as related to net income changes is a convenient way of accounting for the free-rider problem associated with voting on tariff policies. On the problem of free riding and political participation, see Mancur Olson (1965).

decrease is denoted by $0 \leq \sigma(e) < 1$, where σ is positively related to the left-hand expression in (17). The total number of votes cast for a tariff decrease, A, is

$$(18) \qquad A = L\int_0^{\tilde{\tilde{e}}}\sigma(e)f[k(e)]\,de,$$

where $\tilde{\tilde{e}}$ is the factor-ownership index for persons with zero net gains from the tariff decline; that is, for those with $D[k(\tilde{\tilde{e}}),t] = c(\tilde{\tilde{e}})$.

If, at a given tariff rate, $E > A$, a majority will vote for higher tariffs. As the tariff is raised, the voting strength of proponents (E) for a tariff increase will decline while opponents (A) increase their votes. This follows from (16) and (18), as σ, $\tilde{e}$, and $\tilde{\tilde{e}}$ are directly and ρ is inversely related to t. An equilibrium tariff finally is established, when $E = A$, or[15]

(19)

$$\int_{\tilde{e}}^1 \rho(e)f[k(e)]\,de = \int_0^{\tilde{\tilde{e}}}\sigma(e)f[k(e)]\,de.$$

It is clear that the equilibrium tariff depends on the factor distribution at both tails, that is, for people with indexes $e \geqq \tilde{e}$ and $e \leqq \tilde{\tilde{e}}$, as well as on voting probabilities on the part of winners, $\rho(e)$, and losers, $\sigma(e)$. Whether a person is going to vote at all and with what probability, in turn, depends on the person's factor-ownership ratio, voting costs, and actual tariff rate.

It is relatively easy to see that the equilibrium tariff rate with voting costs, t^*, is less than the optimal tariff rate of the marginal gainer from a tariff increase, $\tilde{t}(\tilde{e})$, and larger than the optimal tariff rate of the marginal gainer from a tariff decrease, $\tilde{t}(\tilde{\tilde{e}})$.[16]

Hence, the equilibrium tariff will be the optimal tariff of a nonvoting factor owner.

Depending on the specifically made assumptions on factor-ownership distributions and voting probabilities, different hypotheses on tariff determination can be developed. Here, it only should be noted that free trade would be adopted if factor-ownership distribution, voting costs, and voting probabilities were symmetric.[17]

Earlier I mentioned the role of voter eligibility restrictions as a possible determinant of actual tariff rates. Even today, every country excludes some factor owners from voting. Major restrictions result from age, residency, and citizenship criteria. Young factor owners, as well as migrant and alien workers, are generally not eligible to vote. In the past, women and poor people, including small property owners and small taxpayers, were excluded as well. Such eligibility restrictions are really a special form of prohibitive voter participation costs. Their major effect is that the factor-ownership distribution of voters is no longer identical to that of the population. In most instances these restrictions tend to cut off, or at least flatten, the left tail of the factor-owner distribution. With the exclusion of many "capital poor" people, eligible voters are more likely to protect the interests of capital owners; if the factor-ownership distribution of all people is symmetric, a tariff on the capital-intensive import good will be voted in as the factor-ownership ratio of the median voter exceeds the endowment ratio; and even if there is positive skewness, as suggested before, it still is possible that the median voter's optimal tariff is positive.

Changing voting rules affects the identity of the median voter and thereby tariff policy. If it is accepted that past voting restrictions were directed against workers and small capital owners, one should observe that the gradual dismantling of voting restrictions has been associated with or followed by declining tariffs on capital-intensive imports and rising tariffs on labor-intensive imports. Em-

[15]Adjustment to such an equilibrium implies the possibility of repeated voting on the same tariff issue. While repeated voting in the sense of going again and again to the voting booth is quite unrealistic, repeated voting through other forms of communications from the public to the politician is far more realistic.

[16]At the equilibrium tariff, t^*, there exists a marginal voter among net gainers from a tariff increase for whom $B[k(\tilde{e}),t^*] = c(\tilde{e})$. Without voting costs, the marginal gainer's optimal tariff $\tilde{t}(\tilde{e})$ is such that $B[k(\tilde{e}),t] = 0$. Since $(\partial B/\partial t) < 0$ and $c(\tilde{e}) > 0$, it must be that $t^* < \tilde{t}(\tilde{e})$. Similarly, one can show that $t^* > \tilde{t}(\tilde{\tilde{e}})$.

[17]Symmetric voting probabilities mean that, for all $(B-c) > 0$, $\rho = \sigma$ if $(B-c) = (D-c)$.

pirical support for this hypothesis, as presented by Bennett Baak and Edward Ray (1982) for the United States, is not fully conclusive.

III. Protection of Industry-Specific Interests[18]

The discussion of tariff formation in a Heckscher-Ohlin type economy emphasized the role of factor-ownership distributions, voting costs, and voter eligibility in determining tariffs. Where a given factor is employed (i.e., its association with a given industry) was irrelevant. This framework seems most suitable in explaining long-run tariff trends, especially in relationship to changes in voter eligibility rules, voting costs, or overall factor-ownership distributions. It is much less suitable, however, in explaining more short-term attempts by individual industries to gain tariff protection. In particular, it sheds no light on the frequently observed phenomenon that a single industry succeeds in raising tariffs on its product, even though the vast majority of eligible voters does not benefit from such a policy. In the following, I introduce a many-industries model with specific factors to show how majority voting can result in tariff protection of a small industry. First, I lay out the model and discuss the impact of a tariff increase on a person's income share. Later, I demonstrate how a small minority of gaining factor owners can become a majority of actual voters for a tariff increase on a given commodity.

A. *The Model*

The production structure of this section is described by Jones' (1975) multicommodity model with specific factors. There are $n > 2$ industries, each of which employs a fixed amount of a factor specific to that industry, V_j, $j = 1,\dots,n$, and a variable amount of a mobile factor, V_N. The mobile factor is employed in all industries. Since we are again concerned with the implications of tariffs for income shares of different individuals, the main interest lies in the relationships between commodity and factor prices in such a model. When needed, I will take these relationships from the Jones paper, using the same notation where appropriate.

With many commodities, the ith individual's indirect utility function is stated as

$$(20)\qquad U^i = U^i\left(p_1, 1, p_3, \dots, p_N; \phi^i Y\right),$$

where p_j is the domestic price of commodity j in terms of the second commodity. The ith person's income share in the presence of n specific factors is

$$(21)\qquad \phi^i = \left[R_N V_N^i + \sum_{j=1}^{n} R_j V_j^i\right] \Bigg/ \left[R_N V_N + \sum_{j=1}^{n} R_j V_j\right],$$

where R_j, $j = 1,\dots,n$, and R_N denote returns on the jth specific and the mobile factor, respectively, V_j^i and V_N^i are the ith person's ownership of the jth specific and mobile factor, and V_j and V_N are the corresponding factor endowments of the country. Let us consider the mobile factor to be labor and assume, as before, that each individual owns one unit; that is, $V_N^i = 1$.

Total income, Y, consists again of domestic factor income and tariff revenues. It is assumed that the first m commodities are imported and the other $(n - m)$ commodities are exported in the relevant range of tariff rates. Denoting imports (exports) of the jth commodity by M_j (E_j) and the corresponding import (export) tariff by t_j (τ_j), the economy's income is

$$(22)\qquad Y = \sum_{j=1}^{n} p_j X_j + \sum_{j=1}^{m} t_j \pi_j M_j + \sum_{j=m+1}^{n} \tau_j p_j E_j.$$

[18]The association of tariffs with the economic interest of sector-specific groups is one of three models tested by Caves.

As was the case in the two-by-two Heckscher-Ohlin model, it is again asked how the ith individual's welfare is affected by a given tariff change, say on the gth commodity. Differentiation of (20) with respect to t_g and some manipulations yield:

$$(23)\quad \left(\partial U^i/\partial t_g\right) = \left(\partial U^i/\partial y^i\right)\left\{Y\left(\partial\phi^i/\partial t_g\right) + \phi^i\left[\sum_{j=1}^{m} t_j\pi_j\left(\partial M_j/\partial t_g\right) + \sum_{m+1}^{n} \tau_j p_j\left(\partial E_j/\partial t_g\right)\right]\right\},$$

indicating that personal welfare effects depend on how a given tariff change affects a person's income share, on the one hand, and the tariff-weighted value of imports plus exports, on the other hand. Note that the direction of the change in the tariff-weighted value of imports and exports is generally not known without further restrictions; but it will be zero under the restriction that initially there is free trade. Corresponding to the analysis in Section II, it would be possible to utilize (23) for all $g=1,\dots n$ and derive an optimal tariff vector for individual i. As shown by Murray Kemp (1969, p. 298) for multicommodity optimal tariff structures of a country, such an optimal tariff vector is one-dimensionally indeterminate. And, without imposing severe restrictions, relatively little can be said about the precise structure of the tariff vector. Fortunately, for the issue under consideration, namely how a single industry can succeed in gaining tariff protection, knowledge of the optimal tariff vector is not crucial.

B. *Tariff-Induced Changes in Income Shares*

An individual associates with the interests of a given industry if his or her specific factor ownership is concentrated in that industry. In order to emphasize this association between an individual's and a given industry's interests in describing voting behavior, assume now that each person possesses at most one *type* of specific factor in addition to one unit of the mobile factor. The income share of individual i with specific factor ownership in industry $h, h=1,\dots,n$, therefore, is defined as

$$(21')\quad \phi_h^i = \left(R_N + R_h V_h^i\right)\Big/\left(R_N V_N + \sum_{j=1}^{n} R_j V_j\right).$$

Differentiation of (21′) with respect to t_g indicates how a tariff increase on commodity g affects the ith individual's income share:

$$(24)\quad \left(1+t_g\right)\partial\phi_h^i/\partial t_g = \left[\frac{R_h V_h^i}{\left(R_N V_N + \sum_{j=1}^{n} R_j V_j\right)}\right]\left[\frac{\left(\hat{R}_h - \hat{R}_N\right)}{\hat{p}_g}\right] - \phi_h^i\left\{\sum_{j=1}^{n}\alpha^j\left[\frac{\left(\hat{R}_j - \hat{R}_N\right)}{\hat{p}_g}\right]\right\},$$

where $\alpha^j = R_j V_j/(R_N V_N + \Sigma_{j=1}^{n} R_j V_j)$ is the jth specific factor's distributive share in national factor income,

$$\sum_{j=1}^{n}\alpha^j = 1-\alpha^N,$$

and $\quad \alpha^N = R_N V_N\Big/\left(R_N V_N + \sum_{j=1}^{n} R_j V_j\right).$

The term in braces on the right-hand side of (24) can be interpreted as the "gth commodity's bias with respect to the mobile factor." Ruffin and Jones (calling the mobile factor labor and denoting its return by w_L instead of our R_N) define "commodity j to be unbiased with respect to labor if the relative change in the wage rate (w_L) brought about by an increase in p_j is precisely the average for the changes in all factor prices" (p. 339). And the average of all factor price changes is obtained by weighting each factor's relative price change by the factor's

share in national income. Hence, one can define commodity g to be biased against, unbiased, or biased towards the mobile factor as

$$(25)\quad b_g = \sum_{j=1}^{n} \alpha^j\left[\left(\hat{R}_j - \hat{R}_N\right)/\hat{p}_g\right]$$

$$= \left[\sum_{j=1}^{n} \alpha^j \hat{R}_j + \alpha_N \hat{R}_N - \hat{R}_N\right]/\hat{p}_g \gtreqless 0.$$

The sign of the first term on the right-hand side of (24) crucially depends on the ith individual's association with a given industry. If an individual's specific factor ownership is in the industry whose tariff rate is raised (i.e., $g = h$), one can show that

$$(26)\quad \left(\hat{R}_g - \hat{R}_N\right)/\hat{p}_g = \sum_{\substack{h=1\\ h\neq g}}^{n} \beta_h/\theta_{gg} > 0,$$

where I utilized equations (13) and (14) of the Jones (1975) paper, $\beta_g > 0$ indicates the relative change in the mobile factor's return as the gth commodity's price rises by 1 percent and $\theta_{gg} = (R_g V_g)/(p_g X_g)$ denotes the cost share of the specific factor in the production of commodity g. Substitution of (26) in (24) reveals that an individual's income share will definitely rise if commodity g is biased towards labor or not biased at all. If, on the other hand, an individual's specific ownership is in an industry whose tariff rate is not changed (i.e., $g \neq h$), then one can show that

$$(27)\quad \left(\hat{R}_h - \hat{R}_N\right)/\hat{p}_g = -\beta_g/\theta_{hh} < 0.$$

Consequently, the individual's income share will definitely decline, if commodity g is biased against labor or not biased at all. Finally, it should be noted that the first term on the right-hand side of (24) is zero for people with only labor but no specific factor ownership. Their income shares rise, remain unchanged, or fall, as commodity g is biased towards, unbiased, or biased against labor.

In the special case where the above-stated bias is zero (i.e., $b_g = 0$ in (25)), Ruffin and Jones (pp. 339–40) prove that β_g also represents the fraction of total factor income generated in industry g; that is, if $b_g = 0$, then

$$(28)\quad \beta_g = p_g X_g \bigg/ \left(R_N V_N + \sum_{j=1}^{n} R_j V_j\right).$$

Furthermore, it should be noted that under all circumstances—whether commodity g is labor biased or not, it must hold that $\sum_{h=1}^{n} \beta_h = 1$.

In the next subsection, the assumption that the commodity on which the tariff is imposed is unbiased towards labor is used. Under this assumption, only the income shares of specific factor owners in the "protected" industry will rise. Specific factor owners in all other industries will suffer a decline in their income shares, and income shares of people who are workers only will not be affected.

C. *Industry Protection under Majority Voting*

Factor owners are assumed to vote their economic interests. Equation (23) above revealed that a given tariff increase affects a person's welfare through two forces: changes in the person's income share and changes in the tariff-weighted value of exports and imports. One could easily construct a scenario such that the latter change is positive and one could conclude that it is, at least, conceivable that a majority of voters gain from a given tariff increase, even though owners of the specific factor in the protected industry are a minority. While this possibility has to be recognized, the intent here is not to adopt a set of unusual assumptions for demonstrating that a small industry can gain tariff protection under majority voting. Instead, the objective is to postulate a rather neutral set of plausible assumptions to substantiate my contention. These assumptions are

1) Initially, free trade is prevailing and only one industry, namely industry g, tries to gain tariff protection;[19]

[19] If several industries try to obtain tariff protection more or less at the same time, the possibility of logrolling must be allowed for. In this case, a unique equilibrium generally will not exist.

2) Commodity g is unbiased with respect to the mobile factor (i.e., $b_g = 0$);

3) There are voter participation costs and these costs are the same for every voter, whether voting for a tariff or a subsidy.

In the presence of voting costs, the decision to participate in voting is based on an individual's assessment of tariff-induced real income changes relative to voting costs. The real income change from leaving free trade and imposing a tariff on commodity g, as experienced by the ith individual with specific factor ownership in industry h, is given by

$$(29)\quad B^i_{hg} = Y\left(\partial\phi^i_h/\partial t_g\right)$$

$$= Y\left[\left(R_h V^i_h\right)/\left(R_N V_N + \sum_{j=1}^{n} R_j V_j\right)\right] \times\left[\left(\hat{R}_h - \hat{R}_N\right)/\hat{p}_g\right],$$

where the assumptions above and equations (23) and (24) were employed. If individual i owns a specific factor in the industry where the tariff increase takes place, then

$$(30)\qquad B^i_{gg} = Y\left(1-\beta_g\right)\beta_g \lambda^i_g > 0,$$

where $\lambda^i_g = (V^i_g/V_g)$ denotes the ith individual's ownership share of specific factor g, and where we used (26) and (28), the property $\Sigma^n_{h=1;\ h\neq g}\beta_h = (1-\beta_g)$, and the definition of θ_{gg} in reducing (29) to (30). Alternatively, if individual i owns a specific factor employed in any other, not protected industry h, then

$$(31)\qquad B^i_{hg} = -Y\beta_h\beta_g\lambda^i_h < 0,$$

using (27). Finally, for all those people who do not own any specific factor, as $V^i_g = V^i_h = 0$, there is no real income change in the small neighborhood of the initial free-trade equilibrium.[20]

An assessment of real income changes of people owning specific factors in either the protected or one of the many other industries reveals that the former will gain while the latter will lose. Accordingly, only a small fraction of all factor owners would favor such a tariff increase and it would not be voted in by a majority if all factor owners participated in voting. In fact, the majority of people would gain from a subsidy on importing commodity g. In the presence of voting costs, however, the *magnitude of gains or losses* relative to voting costs must be considered as well. Equations (30) and (31) state that the magnitude of real income changes depends on the value of national income, Y, and a person's specific factor-ownership share, λ^i_h. But, most important, it depends on the relative sizes of the protected industry and of the industry whose specific factor is owned, as β_g and β_h measure the respective industries' shares in national income. The industry-size element in affecting real income is most significant since it introduces an asymmetry between the magnitude of gains and losses for people with identical specific factor-ownership shares held in the protected and nonprotected sectors, respectively. More precisely, for all $\lambda^i_g = \lambda^i_h$, where $h = 1,\ldots,n$ and $h \neq g$,

$$(32)\qquad B^i_{gg} > -B^i_{hg},$$

as $(1-\beta_g) > \beta_h$ always, if there are more than two industries. As shown next, this asymmetry between the size of gains and losses implies that, with voting costs, a minority of factor owners may become a majority of participating voters.[21]

Per capita voting costs c were assumed to be the same for both supporters and opponents of the tariff increase. A factor owner definitely will refrain from voting if real income gains from a tariff or subsidy increase do not exceed voting costs. Hence, a person will not participate in voting for a tariff

[20] This is the same conclusion as the one obtained by Ruffin and Jones for the case when tariff proceeds are redistributed (see their p. 344).

[21] One can show that, in the small neighborhood of the initial free-trade point, the sum of total real income gains is exactly equal to the sum of real income losses when all factor owners are accounted for.

increase if

$$B^i_{gg} = Y\beta_g(1-\beta_g)\lambda^i_g \leqq c. \tag{33a}$$

And a person will not participate in voting for a tariff decrease or subsidy if

$$D^i_{hg} = -B^i_{hg} = Y\beta_g\beta_h\lambda^i_h \leqq c, \tag{33b}$$

where D^i_{hg} is the subsidy-induced real income gain to person i owning some of specific factor h. People with factor ownership shares $\lambda^i_g \leqq \tilde{\lambda}^i_g$ and $\lambda^i_h \leqq \tilde{\tilde{\lambda}}^i_h$ definitely will not participate in voting, where

$$\tilde{\lambda}^i_g = c/\left[Y\beta_g(1-\beta_g)\right] \tag{34}$$

$$\text{and} \quad \tilde{\tilde{\lambda}}^i_h = c/\left[Y\beta_g\beta_h\right], \quad h \neq g;\ h = 1,\dots,n,$$

can be called the specific factor-ownership share of the marginal voter for a tariff and subsidy, respectively. Comparing these ownership shares of marginal nonparticipants in voting, one can see that

$$\tilde{\tilde{\lambda}}^i_h > \tilde{\lambda}^i_g \tag{35}$$

as $[(1-\beta_g)/\beta_h] > 1$ always. The separation between definite nonvoters and potential voters occurs at a much higher endowment share in the many nonprotected industries than in the single protected industry. Hence, it is quite possible that only a few large specific factor owners in each losing industry remain potential resisters to the tariff increase, while a large number of factor owners in the gaining industry tend to support it. In fact, it is quite possible that losers will pose no active resistance at all to the proposed departure from free trade.

A simple example will illustrate the magnitudes by which the marginal nonparticipants' factor-ownership ratios may differ between gainers and losers. First, assume there are 10 industries: for each industry $\beta_h = .1$, and voting costs and income are such that $\tilde{\lambda}^i_g = 1/10{,}000$. Then $(1-\beta_g)/\beta_h = 9$ and $\tilde{\tilde{\lambda}}^i_h = 9/10{,}000$. Second, assume there are 100 industries $\beta_h = .01$ for each industry, and again $\tilde{\lambda}^i_g = 1/10{,}000$. In this case, $\tilde{\tilde{\lambda}}^i_h = 99/10{,}000$. The examples illustrate that the marginal nonparticipant's factor-ownership share is far greater among losers than gainers, and that it tends to rise the more diversified the economy is (i.e., the more sectors it has).

So far, I have only tried to describe the separation line between definite nonvoters and potential voters among factor owners in both protected and nonprotected industries. The number of potential voters in the two opposing camps crucially depends on the height of voting costs, the distribution of specific factor ownerships in the various sectors, and the degree of diversification of the economy. As mentioned above, it is quite possible that a minority of factor owners gaining from a tariff increase becomes a majority of potential voters. However, even if losing factor owners maintain a majority of potential voters, it still is quite possible that the gaining industry's factor owners will succeed in obtaining tariff protection. What matters for tariff policy formation are actual votes rather than potential votes, and the minority of potential votes among gainers could easily become a majority of actual votes as will be shown next.

The probabilities of potential voters' participation in voting depend, as mentioned in Section II, on the magnitude of net gains from a certain outcome. Denoting the probability of a specific factor owner in industry g to vote for the tariff increase by ρ_g and the probability of a specific factor owner in industry h to vote for a subsidy or tariff decrease by σ_h, we have

$$\rho_g = \rho\left[B_{gg} - c\right] \text{ and } \sigma_h = \sigma\left[D_{hg} - c\right], \tag{36}$$

where $\rho_g = 0$ for $(B_{gg} - c) \leqq 0$, $\sigma_h = 0$ for $(D_{hg} - c) \leqq 0$ and $\rho'_g > 0$, $\sigma'_h > 0$ in the range of positive net gains. Unrestricted comparisons of ρ_g with σ_h are, of course, not possible. If, however, it is assumed that the voting probabilities for given values of net gains are the same, then we can compare the voting probabilities of factor owners in different industries with the same specific factor ownership shares. Assuming that $\lambda_g = \lambda_h$, one can see that

$$\begin{aligned}\rho_g = \rho\left[B_{gg} - c\right] &> \sigma_h \\ &= \sigma\left[\left(B_{gg} - c\right) - \left(B_{gg} - D_{hg}\right)\right],\end{aligned} \tag{37}$$

since $B_{gg} > D_{hg}$, using (32) and (33b). Potential voters with identical factor ownership shares are more likely to vote if they are gainers than if they are losers from a tariff increase. Hence, it is quite possible that a majority of potential voters against a tariff increase becomes a minority of actual voters once voting probabilities are accounted for.

I have demonstrated that the existence of voting costs can be most instrumental in helping a small industry gain tariff protection. The push for protection is, however, not unlimited since resistance to tariff increases will rise and the forces for tariff increases will weaken as the tariff rate, t_g, further departs from zero. This can be seen from (23) which for $t_g > 0$ and all other $t_j = \tau_j = 0$, becomes

$$(23') \quad \left(\partial U^i/\partial t_g\right) = \left(\partial U^i/\partial y^i\right)\left\{Y\left(\partial \phi^i/\partial t_g\right) + \phi^i t_g \pi_g\left(\partial M_g/\partial t_g\right)\right.$$

Since $(\partial M_g/\partial t_g) < 0$, each person's real income change is affected by a negative force in addition to the income share change. The higher t_g, the less the gainers will gain and the more the losers will lose. Hence, the drive for a further increase in t_g will eventually stall as no majority can be formed to keep raising the tariff rate.

My findings support Baldwin's conjecture that a small group of big gainers may succeed in obtaining tariff protection. Crucial for this occurrence is that each person's specific factor ownership is relatively concentrated in one industry and that there are voting costs. In this case, a large number of losers from a tariff increase becomes a much smaller number of potential voters which, in turn, becomes an even smaller number of actual voters. Among gaining factor owners, there also will not be full voter participation, but the shrinkage tends to be much less. In the end, the gainers may be successful in gaining tariff protection through a majority vote.

V. Concluding Remarks

This paper argues that a country's actual tariff policy is the result of its underlying factor-ownership distribution. There are two links in the chain of causality between factor-ownership distribution and tariff policy. The first link describes how the economic interests of people in a given tariff policy are related to the factor-ownership distribution. One may call it the "economic link" since the economy's production structure is the main determinant of the process through which a person's real income is affected. I deal with two alternative production structures. The two-by-two Heckscher-Ohlin model, that is highly aggregated and frequently associated with long-run portrayals of production structures, seems most useful in explaining long-run changes in the overall tariff structure. The many-commodity model, with specific factors, on the other hand, seems more appropriate in explaining day-to-day attempts by individual industries or interest groups in gaining tariff protection. Irrespective of the production structure, each person is shown to have an optimal tariff policy that generally differs from the actual policy. The second link between factor-ownership distribution and tariff policy may be called the "political link" as it refers to the political process through which economic interests are translated into actual tariff policy. My analysis emphasizes the sensitivity of tariff policies to changing voter eligibility rules and voter participation costs under majority voting. In particular, it is demonstrated how a small minority of factor owners can succeed in gaining tariff protection for its industry under majority voting provided voting costs are significant.

This paper's assumptions concerning both the economic structure and political process were deliberately chosen to allow tracing of the path from factor-ownership distributions to tariff policies. Although the assumptions are quite standard in the trade-theoretic and public choice literature, it would be most desirable to relax some of the restrictions. Concerning the economic link, future research should explore more general descriptions of factor-ownership distributions, imperfect information of factor owners on real income effects of tariff changes, and nonhomothetic tastes. Concerning the political link, one should try to account for multiple-issue voting and logrolling, policy

formation in a representative democracy, and the working of a multiparty system.

REFERENCES

Baak, Bennett D. and Ray, Edward John, "The Political Economy of Tariff Policy: A Case Study of the United States," mimeo., Ohio State University, February 1982.

Baldwin, Robert E., "The Political Economy of Postwar U.S. Trade Policy," *The Bulletin*, Graduate School of Business Administration, New York University, 1976.

______, "The Political Economy of Protection," in J. N. Bhagwati and T. N. Srinivasan, eds., *Import Competition and Response*, Chicago: University of Chicago Press, 1982, 263–86.

Bhagwati, Jagdish N., "Shifting Comparative Advantage, Protectionist Demands, and Policy Response," in his and T. N. Srinivasan, eds., *Import Competition and Response*, Chicago: University of Chicago Press, 1982, 153–84.

Black, Duncan, (1948a) "On the Rationale of Group Decision Making," *Journal of Political Economy*, February 1948, *56*, 23–34.

______, (1948b) "The Decision of a Committee Using a Special Majority," *Econometrica*, July 1948, *16*, 245–61.

Brock, William A. and Magee, Stephen P., "The Economics of Special Interest Politics: The Case of the Tariff," *American Economic Review Proceedings*, May 1978, *68*, 246–50.

______ **and** ______, "Tariff Formation in a Democracy," in J. Black and B. Hindley, eds., *Current Issues in International Commercial Policy and Diplomacy*, London: Macmillan, 1980, 1–9.

Buchanan, James M. and Tullock, Gordon, *The Calculus of Consent*, Ann Arbor: University of Michigan Press, 1962.

Caves, Richard E., "Economic Models of Political Choice: Canada's Tariff Structure," *Canadian Journal of Economics*, May 1976, *9*, 278–300.

Dinopoulos, Elias, "Import Competition, International Factor Mobility and Lobbying Response: The Schumpeterian Industry Case," *Journal of International Economics*, May 1983, *14*, 395–410.

Downs, Anthony, *An Economic Theory of Democracy*, New York: Harper and Row, 1957.

Findlay, Ronald and Wellisz, Stanislaw, "Endogenous Tariffs, The Political Economy of Trade Restrictions, and Welfare," in J. N. Bhagwati and T. N. Srinivasan, eds., *Import Competition and Response*, Chicago: University of Chicago Press, 1982, 223–34.

Helleiner, G. K., "The Political Economy of Canada's Tariff Structure: An Alternative Model," *Canadian Journal of Economics*, May 1977, *10*, 318–26.

Hillman, Arye L., "Declining Industries and Political-Support Protectionist Motives," *American Economic Review*, December 1982, *72*, 1180–87.

Jones, Ronald W., "Tariffs and Trade in General Equilibrium: Comment," *American Economic Review*, June 1969, *59*, 418–24.

______, "Income Distribution and Effective Protection in a Multicommodity Trade Model," *Journal of Economic Theory*, August 1975, *11*, 1–15.

Kemp, Murray C., *The Pure Theory of International Trade and Investment*, Englewood Cliffs: Prentice Hall, 1969.

Marvel, Howard P. and Ray, Edward J., "The Kennedy Round: Evidence on the Regulation of International Trade in the United States," *American Economic Review*, March 1983, *73*, 190–97.

Mueller, Dennis C., *Public Choice*, Cambridge: Cambridge University Press, 1979.

Olson, Mancur, *The Logic of Collective Action*, Cambridge: Harvard University Press, 1965.

Peltzman, Sam, "Toward a More General Theory of Regulations," *Journal of Law and Economics*, August 1976, *19*, 211–40.

Pincus, Jonathan J., "Pressure Groups and the Pattern of Tariffs," *Journal of Political Economy*, August 1975, *83*, 757–78.

Ray, Edward J., "The Optimum Commodity Tariff Rates in Developed and Less Developed Countries," *Review of Economics and Statistics*, August 1974, *56*, 369–77.

Riker, William H. and Ordeshook, Peter C., "A Theory of the Calculus of Voting," *American Political Science Review*, March 1968, *62*, 25–42.

Ruffin, Roy and Jones, Ronald W., "Protection and Real Wages: The Neoclassical Am-

biguity," *Journal of Economic Theory*, April 1977, *14*, 337–48.

Wellisz, Stanislaw and Findlay, Ronald, "Some Thoughts on Protection and Rent-Seeking in Developing Countries," International Economics Research Center Paper No. 30, Columbia University, June 1983.

______ **and Wilson, John D.,** "A Theory of Tariff Formation," mimeo., Columbia University, August 1983.

Young, Leslie and Magee, Stephen P., "Factor Returns and Resource Allocation in the Political Economy of Trade Distortions," presented at the North American Summer Meeting of the Econometric Society, Northwestern University, June 1983.

Consensus and Dissension Among Economists: An Empirical Inquiry

By BRUNO S. FREY, WERNER W. POMMEREHNE, FRIEDRICH SCHNEIDER, AND GUY GILBERT*

Among its practitioners, there is a widespread feeling that modern economics has established itself as a science enjoying a high degree of consensus. Jürg Niehans for instance, states that "... there is, at any given time, a large amount of consensus. In fact, the economic doctrines taught in Nairobi, São Paulo, New Delhi, Tel Aviv, Geneva, Harvard, and Bowling Green State... have a large common core" (1981, p. 170). Other signs, however, indicate that there is much less unanimity in *practical economic policy*. The *Herald Tribune* for example, writes "Almost nobody has a good word for economists any more, including economists," and goes on to tell the story of President Truman, who after receiving so much advice in the form of "on the one hand... but on the other hand" expressed the wish for a "one-handed economist" (International Edition, Jan. 25, 1982).

What economists think, and whether there is consensus among economists, would not be a matter of concern if *beliefs* do not have a very strong effect on economic policy decisions and on the state of the economy. In an analysis of inflation, Martin Feldstein, for example, states: "It was not events but *ideas* that propelled the increasing rate of inflation" (1982, p. 63), and "the upward drift of the inflation rate was the result of a fundamental set of beliefs about the economy and about macroeconomic policy that was shared by economists and policy officials during the past two decades" (pp. 63–64).

We seek to determine the degree to which certain beliefs are widely shared among economists in a given country and across different countries. The present study is based on extensive surveys made in the United States and in four European countries: France, the Federal Republic of Germany, Austria, and Switzerland.[1]

I. The Survey

A stratified random sample of 2,072 economists with respect to country, position, and occupation was asked to respond to 27 propositions about economic problems. They could "generally agree," "agree with provisions," or "generally disagree," or they could, of course, refuse to answer any specific assertion or just not send back the questionnaire. The economists were (randomly) chosen from a list of the members of professional associations of the respective countries.[2] Complete

*The first three authors are professor and lecturers, respectively, University of Zurich, Kleinstrasse 15, Zurich, Switzerland. The last is lecturer, University of Paris-X-Nanterre, 200 Avenue de la République, 92001 Nanterre Cedex, France. Earlier versions of this paper have been critically reviewed by a number of economists in various countries. We are particularly grateful to Peter Bernholz (Basel), Mark Blaug (London), Bernard Bobe (Paris), Holger Bonus (Konstanz), Knut Borchardt (Munich), Pierre-Henri Derycke (Paris), René L. Frey (Basel), Herbert G. Grubel (Vancouver), Gebhard Kirchgaessner (Zurich), Jean-Dominique Lafay (Poitiers), Mico Loretan (Zurich), Fritz Neumark (Frankfort), Andrew Reschovsky (Boston), Kurt W. Rothschild (Linz), Hannelore Weck-Hannemann (Zurich), Burton A. Weisbrod (Madison), Carl-Christian von Weizsäcker (Berne), and Alain Wolfelsperger (Paris).

[1]The country-specific results are published for the United States by James Kearl et al. (1979), for France by Bernard Bobe and Alain Etchegoyen (1981), for the Federal Republic of Germany by Schneider, Pommerehne, and Frey (1983), for Austria by Pommerehne, Schneider, and Frey (1983), and for Switzerland by Frey et al. (1982). (A survey of the opinions of British Members of Parliament based on a more restricted set of questions has been undertaken by Samuel Brittan, 1973.)

[2]In the United States, the American Economic Association; in France, the Association Nationale des Docteurs en Sciences Economiques, the Association Française de Science Economique, and the Centre Na-

replies were returned by 936, and were used for our study. The return rate of 45.2 percent is reasonably good for a study of this kind, particularly as the anonymity of the survey prevented the use of a reminder. The propositions and response categories are derived from the survey for the United States, the results of which are used in our study (as well as those for France).[3] The surveys for Austria, Germany (FRG), and Switzerland were conducted by us.

Section II analyzes the extent to which there exists a common core of economic doctrines among the whole group of economists, and in which areas there is disagreement. Section III considers the question of whether the views of economists differ significantly among the various countries. It turns out that the American, German, and Swiss economists tend to support the market economy and competition, and therewith typically neoclassical propositions, while the French and Austrian economists are more inclined to support government interventions into the economy.

II. Where is the Consensus among Economists?

In order to establish how much consensus there is among the 936 economists whose replies were usable, two criteria were employed. The first criterion is the standardized relative entropy ρ,[4] which condenses the information contained in the four response categories into one measure. This information-theoretic measure given in Table 1 for the total sample equals zero if there is perfect consensus, and equals one if there is no consensus at all. A low ρ means that it would be redundant to ask many economists (as they have similar views), while a high ρ means that the economists' views are entropic, that is, their answers have little structure. The relative entropy measure is nonlinear, that is, a value of, for example, 0.5 should not be interpreted to lie in the "middle" between perfect consensus and complete dissension.

The second criterion used compares the extreme answers along the support-rejection dimension, that is, those "generally agreeing" and those "generally disagreeing" with a proposition. The propositions and answers for all countries combined are presented in Table 1. Each of the 27 propositions shows the percentage distribution of the responses given by the economists who answered the survey.

Most of the propositions with the highest degree of consensus among the complete set of economists pertain to one issue: the price system is considered to be an effective and desirable social choice mechanism. Accordingly, interventions by the government in its functioning are generally rejected. Wage-price controls (Proposition 14), tariffs and import quotas (1), rent ceilings (15), and an employment guarantee (2) are viewed as harming the economy, while flexible exchange rates (5) are accepted as desirable. There is also agreement that it is not the big oil companies that have driven up the oil price (23). Among the propositions forming the common core among the economists, there is only one which does not fit this picture. Economists do not think that interfering with the price system through consumer protection reduces efficiency (26). Antitrust policy (11) is, of course, of quite a different nature. This type of government intervention is designed to protect the working of the competitive price system. Besides the support of the price system, it is possible to discern another (but much less visible) dimension of economists' views: that the budget should be balanced over the business cycle and not yearly (22) can be interpreted as supporting a Keynesian proposition. A

tional de la Recherche Scientifique; in Germany and Austria, the Gesellschaft fuer Wirtschafts- und Sozialwissenschaften, Verein fuer Socialpolitik, in Austria, furthermore, the Nationaloekonomische Gesellschaft; in Switzerland, the Schweizerische Gesellschaft fuer Statistik und Volkswirtschaft.

[3] We are grateful to the authors of the American and French studies for making the results of their surveys available to us.

[4] We define ρ as the actual entropy divided by the maximum possible entropy over the four response categories (including "no answer"). Entropy is the sum of the probability p_i of a particular response category i multiplied with the natural logarithm of the probability p_i, i.e., $\Sigma p_i \cdot \ln(p_i)$.

TABLE 1—PROPOSITIONS AND RESPONSES, TOTAL SAMPLE OF NATIONS

Propositions	Responses (percent)		Relative Entropy (ρ)
1. Tariffs and import quotas reduce general economic welfare	A	57.0	
	B	30.8	0.72
	C	10.3	
2. The government should be an employer of last resort and initiate a guaranteed job program	A	18.8	
	B	29.9	0.80
	C	48.6	
3. The money supply is a more important target than interest rates for monetary policy	A	45.5	
	B	29.2	0.85
	C	21.3	
4. Cash payments are superior to transfers-in-kind	A	48.3	
	B	27.0	0.87
	C	17.4	
5. Flexible exchange rates offer an effective international monetary arrangement	A	48.0	
	B	36.0	0.78
	C	13.5	
6. The "Corporate State" as depicted by Galbraith accurately describes the context and structure of advanced economies	A	28.7	
	B	42.1	0.84
	C	26.8	
7. A minimum wage increases unemployment among young and unskilled workers	A	41.3	
	B	25.5	0.83
	C	31.5	
8. The government should index the income tax rate structure for inflation	A	33.0	
	B	25.2	0.84
	C	39.6	
9. Fiscal policy has a significant stimulative impact on a less than fully employed economy	A	46.8	
	B	39.6	0.78
	C	9.2	
10. The distribution of income in the developed industrial nations should be more equal	A	33.8	
	B	31.6	0.86
	C	31.7	
11. Antitrust laws should be used vigorously to reduce monopoly power from its current level	A	48.9	
	B	37.3	0.75
	C	12.5	
12. Inflation is primarily a monetary phenomenon	A	22.8	
	B	28.8	0.83
	C	45.9	
13. The government should restructure the welfare system along lines of a "negative income tax"	A	28.0	
	B	28.9	0.89
	C	38.1	
14. Wage-price controls should be used to control inflation	A	9.8	
	B	20.7	0.62
	C	68.5	
15. A ceiling on rents reduces the quantity and quality of housing available	A	56.1	
	B	27.2	0.74
	C	15.0	
16. The central bank should be instructed to increase the money supply at a fixed rate	A	15.2	
	B	28.7	0.79
	C	52.6	
17. Effluent taxes represent a better approach to pollution control than imposition of pollution ceilings	A	32.0	
	B	29.3	0.87
	C	35.3	
18. The government should issue an inflation indexed security	A	23.3	
	B	24.1	0.83
	C	48.9	
19. The level of government spending should be reduced (disregarding expenditures for stabilization)	A	39.4	
	B	27.8	0.85
	C	30.3	

TABLE 1—CONTINUED

Propositions	Responses (percent)		Relative Entropy (ρ)
20. The central bank has the capacity to achieve a constant rate of growth of the money supply if it is so desired	A	23.8	
	B	45.1	0.84
	C	27.8	
21. Reducing the influence of regulatory authorities (e.g., in air traffic) would improve the efficiency of the economy	A	29.9	
	B	33.9	0.87
	C	33.0	
22. The federal budget should be balanced over the business cycle rather than yearly	A	54.5	
	B	27.5	0.77
	C	15.5	
23. The fundamental cause of the rise in oil prices of the past five years is the monopoly power of the large oil companies	A	9.3	
	B	21.1	0.64
	C	67.8	
24. The redistribution of income in the developed industrial nations is a legitimate task for the government	A	46.9	
	B	32.4	0.80
	C	19.1	
25. In the short run, unemployment can be reduced by increasing the rate of inflation	A	28.6	
	B	34.0	0.86
	C	34.8	
26. "Consumer protection" laws generally reduce economic efficiency	A	13.6	
	B	23.6	0.70
	C	61.4	
27. The economic power of labor unions should be significantly curtailed	A	22.0	
	B	31.0	0.82
	C	44.8	

Notes: $N = 936$. A = Generally Agree; B = Agree with Provisions; C = Generally Disagree. The responses do not sum to 100 percent as the "No Answer" category is not listed.

majority of economists thinks, however, that the supply of money (rather than interest rates) should be controlled (3) which many would regard as a monetarist proposition. Nevertheless, there is much less consensus (or even marked dissension) on other monetarist propositions such as Propositions 12, 16, or 20.

The propositions about which there is most dissension among the 936 economists, according to the entropy measure, may be arranged along four dimensions. As may be expected, there is a considerable spectrum of opinions about propositions dealing with issues which are in the center of recent political debate: regulation (21), Phillips curve (25), government spending (19). The same applies to "classical" areas of disagreement between the political "right" and "left": distribution of income (10), and corporate state (6). Surprisingly, there is relatively little consensus about basic neoclassical propositions found in most textbooks: cash vs. in-kind transfers (4), effluent taxes (17), and negative income tax (13). As already mentioned, monetarism (in particular, Propositions 3 and 20) is another area of dispute among economists.

Are economists more likely to agree on positive ("can") questions than on normative ("should") questions? Are they more likely to agree on microeconomic than on macroeconomic propositions? To answer these questions, James Kearl et al. (1979, p. 33) classified the propositions in the following way (our proposition numbers follow in parentheses):

Micro "Can": Tariffs (1), cash vs. in-kind transfers (4), flexible exchange rate (5), minimum wage (7), rent ceiling (15), effluent taxes (17);

Micro "Should": Antitrust laws (11), regulation and efficiency (21), consumer protection (26), union power (27);

Macro "Can": Money vs. interest-rate target (3), fiscal policy stimulus (9), inflation as monetary phenomenon (12), money rule is achievable (20), Phillips curve (25);

Macro "Should": Employer of last resort (2), indexed taxes (8), negative income tax (13), pursue money rule (16), indexed securities (18), budget balance (22).

On the basis of a 2×2 analysis of variance of the relative entropy measures, Kearl et al. found for the American economists that there is a significantly higher degree of consensus on positive than on normative questions, and on micro than on macro propositions. These questions are now addressed on the views of the more than 900 economists of the countries included in our study. The average value of the relative entropy ρ is 0.80 for micro and 0.83 for macro questions, and 0.82 for positive and 0.81 for normative propositions, respectively. Undertaking a 2×2 analysis of variance we find for these hypotheses *no* statistically significant difference in the degree of consensus among the propositions listed above (F-values of 2.61 and 0.74, each with 1 and 19 degrees of freedom). The result for the United States does not apply to our international study, neither does it apply to any of the European countries on its own.

III. National Particularities

The preceding section has shown that whereas economists in all countries surveyed, on the whole, tend to support the price system as an effective social decision-making mechanism, there is disagreement concerning other topics, even those involving positive and microeconomic aspects of economics. One cause for this disagreement may be that economists living in the various countries have experienced different historical developments and the traditions are based on different cultural backgrounds.[5] Therefore, in Table 2, the propositions and answers for each individual country are presented.

A striking example of disagreement is provided by Proposition 13 which suggests that a "negative income tax" should be introduced. While 56 percent of the American economists "generally agree" with this proposition, it is supported by less than one-fifth of the respondents in the European countries. The main reason for this large difference may be attributed to the fact that the term "negative income tax" is seldomly used in Europe and that the problems of the welfare system and the attempts to reform it are seen in a quite different light. The distinction, however, does not always lie between American and European economists. Of the French economists, for example, 27 percent "generally disagree" with the proposition that tariffs reduce economic welfare (1), a view which is shared by only between 3 and 13 percent of the economists of the four other countries. The French economists provide several other examples of a position quite different from those of their colleagues in the other countries (for example, Propositions 16, 21, 25).

The national differences in the responses to the propositions are analyzed with the likelihood-ratio text (χ^2, also shown in Table 2) over the frequencies of the four response categories between the countries for each of the 27 propositions. This allows us to answer the question of whether the set of economists in the five countries who answered the surveys may be considered as belonging to one common sample. The high values of the χ^2-statistics for each proposition suggest that the opinions of economists about economic theory and policy do differ between the countries surveyed. This does not, of course, mean that there may not exist groups of countries within which there is a high similarity of opinions. It is interesting to inquire what these national particularities may be, and between which groups of countries they exist. For this purpose, propositions are selected for which the answers of a group of countries deviate most strongly from those of another group of countries (or of one single country), according to the χ^2 test.[6] The analysis will

[5] This also includes contrasting ways of expressing the same topic, leading to different responses to the same sort of proposition.

[6] An alternative approach is to undertake probit analysis taking other influences besides nationality explicitly into account. This has been done in the country studies for Germany, Austria, and Switzerland.

TABLE 2—RESPONSES TO THE PROPOSITIONS, INDIVIDUAL COUNTRIES

Propositions		United States ($N=211$)	Austria ($N=91$)	France ($N=162$)	Germany ($N=273$)	Switzerland ($N=199$)	χ^2
1	A	79.2	43.9	26.5	69.6	47.2	
	B	15.6	41.8	43.9	24.2	40.2	160.8
	C	2.8	13.2	26.5	5.5	10.1	
2	A	25.1	30.8	27.2	8.1	14.6	
	B	26.1	35.1	27.8	27.1	37.2	77.3
	C	45.5	33.0	38.2	63.7	47.2	
3	A	45.0	27.5	42.6	41.0	62.9	
	B	21.8	39.5	26.5	35.2	26.1	62.2
	C	27.5	29.7	23.5	20.9	9.5	
4	A	65.4	42.8	46.9	46.2	36.6	
	B	23.2	35.2	23.5	26.0	31.7	56.7
	C	7.6	18.7	19.1	20.5	21.6	
5	A	60.3	34.1	11.1	62.0	52.3	
	B	33.6	49.4	38.3	30.0	38.7	66.5
	C	4.7	16.5	44.4	5.1	7.5	
6	A	17.1	36.3	32.1	33.7	28.1	
	B	31.7	38.4	43.8	42.5	52.8	149.5
	C	45.0	24.2	22.2	23.1	17.6	
7	A	66.4	29.7	16.7	44.3	35.6	
	B	21.3	34.1	21.0	24.9	30.7	73.2
	C	10.0	35.1	59.8	29.7	32.2	
8	A	38.9	24.2	50.0	23.1	30.7	
	B	25.6	26.4	29.0	21.2	26.6	68.4
	C	30.8	47.2	19.1	53.9	42.2	
9	A	63.5	51.6	54.3	31.1	42.2	
	B	26.5	36.3	37.7	51.3	40.7	46.7
	C	7.6	6.6	4.9	11.7	12.1	
10	A	38.4	46.1	46.3	24.9	25.1	
	B	29.9	30.8	30.9	30.8	35.7	35.2
	C	27.5	19.8	21.6	41.7	36.2	
11	A	47.8	49.4	55.6	54.9	36.7	
	B	35.1	36.3	37.0	34.1	44.7	63.7
	C	14.7	11.0	6.2	10.3	18.6	
12	A	26.1	13.2	10.5	24.5	31.2	
	B	28.9	27.5	18.5	30.8	35.1	171.3
	C	41.2	56.0	67.9	42.9	32.7	
13	A	56.4	22.0	17.9	20.5	19.1	
	B	33.2	27.5	33.3	26.4	24.6	172.1
	C	7.6	42.8	43.2	45.8	53.8	
14	A	5.7	17.6	25.3	2.2	8.5	
	B	21.8	29.7	29.0	5.1	30.2	227.2
	C	71.1	51.6	43.2	92.3	61.3	
15	A	76.8	45.0	21.0	71.8	46.2	
	B	19.4	44.0	30.9	21.2	33.2	90.3
	C	1.9	11.0	43.8	5.9	19.6	
16	A	13.3	5.5	32.7	9.5	15.1	
	B	24.2	24.2	32.7	26.7	35.2	70.2
	C	58.3	68.1	28.4	62.7	44.7	
17	A	47.9	20.9	27.2	34.4	21.1	
	B	29.4	22.0	27.2	29.7	34.2	41.1
	C	18.0	54.9	40.7	33.0	42.7	
18	A	30.8	20.9	33.3	15.4	19.1	
	B	23.2	24.2	25.3	22.7	26.1	45.6
	C	39.4	52.7	37.7	60.1	50.8	
19	A	32.2	42.8	32.1	48.8	38.7	
	B	21.8	25.3	26.5	30.4	32.7	69.0
	C	40.8	30.8	37.1	20.1	27.1	
20	A	33.2	14.3	14.8	18.7	32.7	
	B	38.9	40.6	41.4	49.8	50.2	99.4
	C	22.7	40.7	40.1	30.0	14.1	

(Continued)

TABLE 2—CONTINUED

Propositions		United States ($N = 211$)	Austria ($N = 91$)	France ($N = 162$)	Germany ($N = 273$)	Switzerland ($N = 199$)	χ^2
21	A	45.0	22.0	15.4	30.8	27.6	
	B	29.9	34.1	21.6	44.3	34.2	48.2
	C	20.9	42.8	56.2	23.1	36.2	
22	A	50.7	70.3	44.4	50.3	65.4	
	B	28.5	18.7	30.2	28.9	26.1	46.5
	C	16.1	8.8	19.8	20.1	8.0	
23	A	10.4	16.5	8.6	5.5	10.1	
	B	13.3	28.6	15.4	20.5	31.7	37.6
	C	71.6	53.8	73.5	73.3	58.2	
24	A	50.2	58.2	56.8	40.2	39.2	
	B	28.0	29.7	30.9	38.5	31.2	31.4
	C	18.5	11.0	11.1	20.9	27.6	
25	A	29.4	36.3	19.1	27.8	33.2	
	B	31.3	36.2	34.0	33.0	37.2	62.8
	C	34.1	25.3	42.6	38.8	28.1	
26	A	23.2	6.6	4.9	10.3	18.1	
	B	27.0	22.0	16.7	24.2	25.6	44.8
	C	46.0	70.3	77.2	65.1	55.8	
27	A	30.8	17.6	19.1	20.9	18.6	
	B	37.0	28.6	22.2	34.8	28.1	42.7
	C	28.9	52.7	54.4	43.6	51.3	

Notes: See Table 1; Germany (FRG).

be confined to one issue/dimension of the propositions, namely, the attitude toward the market economy, competition, and government interventions.

In the United States, (mainstream) economists are commonly considered to be strong supporters of the price system and of competition. After World War II, (West) German economists followed suit (see Kurt Rothschild, 1964). In Switzerland, academic economists have been strongly influenced by both postwar American and German economists, and can therefore be counted as belonging to the same group. Accordingly, in all three countries, a tendency to support the market and competition, and to resist government interventions (except in order to maintain competition by antitrust laws) is to be expected.

The opposite is hypothesized for the views of Austrian and French economists. Austria has a long record of government interventions in economic affairs. The Keynesian doctrine is still applied in practical economic policy and finds wide support in the writings of the Austrian economists.[7] France, too, has a long and pronounced record of government involvement into the economy. Economists in Austria and France are thus expected to be rather sceptical about the virtues of the price system and competition, and to tend to support government interventions.

The hypothesis of markedly different opinions between American, German, and Swiss economists on one side, and Austrian and French economists on the other, regarding the market, competition, and government interventions (Propositions 1, 2, 5, 7, 14, 15, 21, 24, 26), tend to be supported by the information contained in Table 3. According to the χ^2-statistics in this table, there is a

[7] We leave it to the reader to compare the "Austrians" as referred to in this paper to the "Austrians" spoken of in a current movement in economic theory.

TABLE 3—LIKELIHOOD-RATIO TEST ON PROPOSITIONS CONCERNING THE MARKET ECONOMY, COMPETITION, AND GOVERNMENT INTERVENTIONS BETWEEN ECONOMISTS OF TWO COUNTRY GROUPS[a]

Propositions	χ^2
1. Tariffs reduce economic welfare	95.1
2. Government as employer of last resort	37.1
5. Flexible exchange rates are effective	173.9
7. Minimum wage increases unemployment	76.7
14. Wage-price controls to fight against inflation	91.2
15. Rent ceiling hurts housing	119.3
21. Reduce the influence of regulatory authorities	60.9
24. Government should redistribute income	33.8
26. Consumer protection laws decrease efficiency	20.4

Note: Proposition text is abbreviated. For all 9 propositions, the *p*-value is smaller than 0.01, that is, the results for the two groups of countries differ significantly at the 99 percent level on a one-tailed-test.

[a]First Group: United States, Federal Republic of Germany, and Switzerland; Second Group: Austria and France.

statistically significant difference between the opinions of the economists of these two groups of countries. Table 2 allows us to undertake a more detailed interpretation of this result.

Between 46 and 64 percent of the American, Swiss, and German as against 33 percent of Austrian and 38 percent of French economists "generally disagree" that "the government should be an employer of last resort" (2). Over 52 percent of U.S., German, and Swiss economists are in favor of flexible exchange rates (5), but only 34 percent of Austrian and 11 percent of French economists are so. Between 70 and 80 percent of the American and German economists generally agree that tariffs reduce economic welfare (Proposition 1); this opinion is shared by only between 27 and 47 percent of economists in the other three countries. The neoclassical "textbook" proposition that "minimum wage increases unemployment among young and unskilled workers" (7) is "generally agreed" to by between 36 and 66 percent of the economists in the United States, Germany, and Switzerland, but by only 30 and 17 percent in Austria and France, respectively. The difference between the "pro-market" and the "pro-government" groups of countries is again conspicuous when it comes to wage-price controls (14). They are generally supported by less than 9 percent of American, German or Swiss economists, but by 18 and 25 percent in Austria and France, respectively. The neoclassical textbook proposition that rent ceilings reduce the stock of housing (15) is accepted by roughly 3 out of 4 American and German economists, but only by 45 percent Austrian and 21 percent French economists (Swiss economists here rather side with the Austrian/French group). The same applies for effluent charges (17) which are rejected by only one-fifth of American and one-third of German economists, but by 41 percent of the French and by more than one-half of the Austrian economists. Between 28 and 45 percent of American, German, and Swiss economists "generally agree" that regulations should be reduced (21); among Austrian or French economists this opinion is held by only 22 and 15 percent, respectively. Government intervention in the form of income redistribution (24) is resisted by 19 to 28 percent of economists in the first group (United States, FRG, Switzerland), compared to only 11 percent in the second group (Austria, France). Finally, less than two-thirds of the economists in Germany and considerably less than one-half in the United States "generally disagree" with statement (26) that "consumer protection laws reduce economic efficiency" whereas the degree of resistance is 70 percent in the case of Austria and 77 percent in the case of France. Overall, the discussion suggests that the American, German, and Swiss economists are clearly more in support of the price system, competition and therewith neoclassical economics, and that the Austrian and French economists are less convinced of the price system and therefore have a higher tendency to agree with interventions of the government into the economy.

IV. Concluding Remarks

The answers given to the 27 propositions by the over 900 economists in five countries

exhibit the highest degree of consensus in one central aspect; namely, that the price system or market is taken to be an effective and desirable social choice mechanism. The propositions about which there is most disagreement were (a) those in which there is an abnormally high frequency of "no answer" responses, which may be attributed to differences in economic policy traditions and unfamiliarity with the terms used; (b) outspokenly normative propositions about income distribution and government spending; and (c) propositions on at the present hotly debated issues such as monetarism or supply-side economics. In general, it could not be confirmed that positive and micro-propositions find a higher degree of consensus than normative and macro-propositions.

The analysis further shows that a major cause for dissension are the differences in views between the economists in the five countries surveyed, attributable to the differences in culture and history as well as to the current economic and political conditions. Economists have had varying experiences with respect to the economic policies practiced in their countries, and therefore have different points of reference. The American, German, and Swiss economists tend to support more strongly the market and competition than their Austrian and French colleagues, who rather tend to view government interventions into the economy more favorably.

REFERENCES

Bobe, Bernard and Etchegoyen, Alain, *Economistes en désordre: consensus et dissensions*, Paris: Economica, 1981.

Brittan, Samuel, *Is There an Economic Consensus?*, London: Macmillan, 1973.

Feldstein, Martin S., "Inflation and the American Economy," *Public Interest*, Spring 1982, *67*, 63–76.

Frey et al., Bruno S., "Welche Ansichten vertreton Schweizer Oekonomen?," *Schweizerische Zeitschrift fuer Volkswirtschaft und Statistik*, March 1982, *118*, 1–40.

Kearl et al., James R., "A Confusion of Economists?," *American Economic Review Proceedings*, May 1979, *69*, 28–37.

Niehans, Jürg, "Economics: History, Doctrine, Science, Art," *Kyklos*, Summer 1981, *34*, 165–77.

Pommerehne, Werner W., Schneider, Friedrich and Frey, Bruno S., "Quot homines, tot sententiae? A Survey among Austrian Economists," *Empirica*, Summer 1983, *15*, 93–127.

Rothschild, Kurt W., "The Old and the New: Some Recent Trends in the Literature of German Economics," *American Economic Review*, March 1964, Suppl., *54*, 2–33.

Schneider, Friedrich, Pommerehne, Werner W. and Frey, Bruno S., "Relata referimus: Ergebnisse und Analyse einer Befragung deutscher Oekonomen," *Zeitschrift fuer die gesamte Staatswissenschaft*, March 1983, *139*, 19–66.

Mortgage Interest Rates in the Populist Era

By BARRY EICHENGREEN*

The high rates of interest from mortgages on land, often exceeding the probable net yield from cultivating the land, have been a familiar feature of many agricultural economies. Usury laws have been directed primarily against encumbrances of this character. And rightly so.

[John Maynard Keynes, 1936, p. 241]

Since the classic work of Solon Justus Buck (1913) on the Granger Movement, historians have attempted to critically assess the economic roots of agrarian discontent at the end of the nineteenth century.[1] The farmers themselves complained that the prices they received for agricultural goods had fallen because railroads and grain elevator operators were acting collusively and middlemen were restricting demand, that the prices they were charged for other commodities were being artificially inflated by suppliers with market power, and that the usurious rates charged by moneylenders on farm mortgages were impoverishing the settler in need of credit. In response, the farmers attempted to organize cooperatives to bypass middlemen and lobbied for the regulation of railroad rates and the imposition of interest rate ceilings. Early analyses of nineteenth-century farm protest, exemplified by John Hicks (1931), while not always taking these complaints at face value, were predicated upon the assumption that farmers were suffering from deteriorating economic conditions.

Subsequent writers, starting with Fred Shannon (1945), attacked the traditional interpretation. Douglass North (1966) provides a summary of the revisionist view. To the complaint that the prices of farm products were falling, he offered that other commodity prices were declining as well and that the farmer's terms of trade were actually improving. To the complaint that railroad rates were artificially inflated, he responded that the price of transportation services fell faster than the general price level, and that the spread between farm prices and market prices narrowed over the period. While admitting that a comparison of mortgage interest rates in the eastern states and the rest of the country was the one observation consonant with the farmer's position, he pointed out that it is hard to know how much of this interest differential was due not to capital market imperfections but to the greater riskiness of mortgage loans out on the frontier (see p. 142).

The subsequent literature went to considerable lengths to elaborate and refine these views.[2] The traditional economic explanations were undermined to the point where textbook descriptions presented agrarian unrest as "The Puzzle of Farm Discontent" (Susan Lee and Peter Passell, 1979, p. 292). Left with no explanation for the frequency with which farmers voiced complaints of distress, economic historians engaged in various attempts to rehabilitate the traditional view. Anne Mayhew (1972) portrayed farm protest

*Department of Economics, Harvard University, Cambridge, MA 02138. An earlier version of this paper was presented to seminars at the University of Rochester and Baruch College. In addition to those made by seminar participants, I am grateful for the comments of Lee Alston, Peter Berck, Stephen DeCanio, Stanley Engerman, Henry Gemery, Robert Higgs, John James, William Parker, Mark Rush, James Stock, Peter Temin, Jeffrey Williams, and Jeffrey Williamson.

[1] In addition to Buck, see the references cited below.

[2] The relevant literature is too extensive to survey here. For examples, see the analysis of railroad rates in Robert Higgs (1970), of farm prices in John Bowman and Richard Keehn (1974), and of agricultural incomes in Robert Fogel and Jack Rutner (1972).

as a response not to declining prices, but to the rising importance of prices themselves—that is, to the commercialization of American agriculture. Joseph Trojanowski (1980) concluded on the basis of a study of Iowa and Kansas that the degree of competition among freight carriers significantly affected the profitability of farming. Robert Klepper (1973) and Robert McGuire (1981) presented evidence for the view that farm protest was a response to economic uncertainty associated with income and price variability. James Stock (1984) argued that fear of foreclosure was the mechanism linking income variability with discontent, and identified a relationship between the level and variability of indebtedness and the location of protest activity.[3]

One strand in the traditional argument that has not been reconsidered is the mortgage interest rate question.[4] There is no disputing the fact that mortgage interest rates tended to be higher on average in western regions of relatively recent settlement than along the eastern seaboard. Neither can one dispute the theoretical possibility that interregional interest differentials simply reflect the risk premium charged by a competitive market for the uncertainties of lending to western farmers in the 1880's and 1890's. Yet, despite the impressive work of Lance Davis (1965), Richard Sylla (1975), John James (1976), and others on the development of American financial markets, no evidence has been presented to test the alternative hypotheses that risk differentials or capital market imperfections account for the observed variation of mortgage rates. Rather, skepticism of the farmers' view is based on the qualitative evidence presented by Allan Bogue (1955) of competitive forces at work, on the suspicion that the Populist characterization of moneylenders as "hyena-faced Shylocks" was overdrawn, and on the aesthetic appeal of the economist's model of competitive forces at work.[5]

What one would like to do is to adjust mortgage interest rates for the risk premium and then to test whether there remains a significant interregional differential in the cost of mortgage credit. In this paper I develop a simple model of mortgage interest rate determination and use regression analysis to estimate its parameters. This enables me to purge the mortgage interest rate of the component associated with risk and then to test for its equality across regions.

The point of departure for the model that follows is the notion that the market prices risk. An acre of farmland is an asset, and its price contains information about the market's assessment of risk and expected returns to farming. However, since the risks of farming differed from the risks of mortgage lending, the purpose of the model will be to provide a link between the two.

The analysis proceeds as follows. Section I describes the growth of mortgage lending in the second half of the nineteenth century. The focus here is on the development of new forms of financial enterprise, on their implications for market structure, and on the statutes adopted by governmental agencies to regulate lending practice. In particular, I discuss the prevalence of statutory interest rate ceilings imposed in the attempt to regulate mortgage lending. Section II presents a model to be used to adjust mortgage interest rates for the risk premium. Ideally, the model would be estimated on a disaggregated data set containing information on the characteristics of individual farms and operators; so in Section III the model is estimated using a special survey of Wisconsin farmers conducted in 1895. Unfortunately, comparable data was gathered only sporadically for other parts of the country. In Section IV, therefore, the same model is estimated using state-level data from the 1890 Census. The results of

[3] P. H. Argersinger (1974) and Jeffrey Williams (1981) use the same measure of indebtedness as does Stock: the ratio of mortgage debt to land value. Argersinger finds that this variable is positively correlated with the Populist vote, but Williams finds the mortgage ratio and the Populists' share to be insignificantly correlated in a multiple regression. This may be due to the limited variability exhibited by this ratio, as discussed below.

[4] See however the discussion in Jeffrey Williamson (1974, pp. 151–53).

[5] See for example Lee and Passell, p. 298; Higgs (1971, pp. 96–97).

the analysis using disaggregated and aggregated data are in broad accord, so in Section V estimates based on state-level data are used to adjust mortgage interest rates for the risk premium. I then use the risk-adjusted rates to test for variations across regions of the country. In particular, I test for interest rate differentials between the North Atlantic states and the remainder of the country, and between states where interest rate ceilings were present and those where no such restrictions were in effect. The results cast new light on the mortgage interest rate question and suggest a new interpretation of the relationship between mortgage interest rates and agrarian unrest.

I. The Growth of Mortgage Lending

Long-term credit played an important role in the development of the American West. In some regions, long-term loans were principally required to finance the costs of "farm-making" rather than the actual purchase of land. With the passage of the Homestead Act of 1862, settlers could claim up to 160 acres of land upon paying minor fees if they agreed to occupy the parcel for five years. In certain states, such as Nebraska, prime land was quickly claimed by railroads, land companies, and the state, and resold at substantial prices (see Bogue, p. 2). But whether the farmer homesteaded his land or purchased it on the open market, funds were needed to obtain timber for buildings and wire for fences. In the plains country, wells had to be dug and coal had to be purchased. To bring his land into cultivation, the farmer needed to obtain working horses, a farm wagon, plows, harnesses, and seeders. In the decade prior to the Civil War, farm-making costs might exceed $1,000, and these capital requirements rose significantly over the course of subsequent decades (see Clarence Danhof, 1941). Few settlers in the second half of the nineteenth century possessed the necessary capital.

Prior to 1860, the western farmer typically borrowed money on the security of his land from local bankers, merchants, and neighbors. (See Ivan Wright, 1923, pp. 100–01.) As these lenders began to accumulate loans, they took to reselling them in the East. Large eastern investors who found the western farm mortgage to be an attractive asset hired agents on the frontier to obtain reliable mortgages exclusively for their eastern client. The farmer typically submitted an application to the agent, who forwarded it to the eastern "home office" along with a critical apppraisal of the borrower and his security. If the loan was approved, funds were sent West by a check drawn on an eastern bank.

By the early 1870's large corporations like the insurance companies of Hartford were investing in farm mortgages in the West (see Bogue, p. 7). Mortgage loan companies increasingly displaced individual middlemen. Once the mortgage companies began to guarantee the mortgages they resold, it was a small step to issue debentures based on a portfolio of mortgages in the manner of modern mutual funds. While the first mortgage companies were organized around 1870, it was not until 1881 that the Iowa Loan and Trust Company of Des Moines issued the first such debentures and until the mid-1880's that the practice became commonplace. (See E. S. Sparks, 1932, pp. 178–79.) By the late 1880's the western farm mortgage had reached great heights of popularity among eastern investors. In 1890 there were at least 167 mortgage companies operating across the United States, and competition among lenders was reported to be intense even in relatively new markets such as Texas.[6]

As agrarian radicalism gained ground in the West over the course of the 1880's, an increasing amount of criticism was directed at the level of interest rates and the lending practices of banks and mortgage companies. Commodity prices had been falling from the mid-1870's, increasing the real burden of the outstanding debt. However, the severity of the problem should not be exaggerated. Prevailing interest rates as well as prices were falling over the period, as they would have had borrowers and lenders operating in a

[6]See D. M. Frederiksen (1894, pp. 212–13); Davis (p. 385); Bogue (p. 160). Other sources provide higher estimates of the number of mortgage companies; see W. F. Mappin (1889, p. 435) and Shannon (p. 189).

TABLE 1—MORTGAGE INTEREST RATES AND WHOLESALE PRICES, 1869–90

Year	Mortgage Interest Rates[a]	Wholesale Prices[b]
1869	7.6	151
1870	8.5	135
1871	8.1	130
1872	8.2	136
1873	8.6	133
1874	8.4	126
1875	7.4	118
1876	8.2	110
1877	7.5	106
1878	7.2	91
1879	7.5	90
1880	7.4	100
1881	7.0	103
1882	6.8	108
1883	6.6	101
1884	6.6	93
1885	6.2	85
1886	6.5	82
1887	6.6	85
1888	6.7	86
1889	6.5	81
1890	6.5	82

Source: Wholesale prices are the Warren and Pearson index for all commodities from U.S. Department of Commerce (1975, Series E-52). Mortgage interest rates from Sylla are unweighted arithmetic averages of rates of earnings on mortgages held by life insurance companies in three regions: New England, the Middle Atlantic States and the Midwest. Sylla's data are from Lester Zartman (1906).

[a]Shown in percent per year.

[b]1910–14 = 100.

competitive market gradually recognized the deflationary trend.[7] (See Table 1.)[8] Moreover, the typical mortgage in the West was five years to maturity, so one might be tempted to think that any increase in the real value of mortgage debt due to a price level declining at an average rate of 1 percent per annum would not be particularly burdensome.[9] Yet the farmers did not see the problem as minor. The price level, rather than declining smoothly, typically fell sharply for short periods, stabilized, and fell again. Bogue reports the difficulties created for the mortgage agent by falling prices and interest rates. Individuals "who had borrowed for five years at 10 percent were clamoring for permission to pay off their loans or at least to be granted a reduction in the rate of interest they were paying. It required much cajoling to keep debtors happy under such conditions" (p. 23).

One response by the farmers was pressure for passage of effective usury laws. By January 1, 1890, thirty-two states and one territory had usury laws in effect. The statutory ceiling on interest charges was 6 percent in eleven states, 8 percent in six states, 10 percent in eleven states, 12 percent in three states and the New Mexico territory, and 18 percent in Idaho. (See 1890 Census, p. 169.) It is by no means certain that these laws were effective. According to the Census Office, the masses of the people believed (ignorantly, in the Office's view) that these laws could be effectively evaded. In eight states usury had no positive penalty, the lender simply being disqualified from collecting the illegal interest if the borrower contested the rate. In only four states and the New Mexico territory was violation of the law a criminal offense, and if conviction was rare, imprisonment was unprecedented. Particularly in the late 1860's and early 1870's it was possible to satisfy the

[7]However, the reaction of nominal interest rates to the rate of deflation was incomplete. On the imperfect adjustment of interest rates on commercial paper, long term bonds, and Treasury bills to changes in the rate of deflation during this period, see Lawrence Summers (1983). The literature contains another explanation for the decline in interest rates in the 1880's, namely that the government's policy of debt retirement reduced interest rates on government securities, which put downward pressure on the entire range of asset returns. See Sylla, ch. V.

[8]See also the time-series for Illinois in R. F. Severson (1962), and for Nebraska, Kansas, and Illinois in Bogue.

[9]According to the 1890 Census, term to maturity averaged 5.36 years in the West but 9.08 years in the nation as a whole. The discrepancy is due largely to an average term of 13.66 years in the Northeast. The five-year figure for the West does not imply that the western debtor freed himself of mortgage obligations within this period, for it was not uncommon for a mortgage to be renewed upon expiration, or for a farmer to take out a second mortgage as a way of paying off a first. Such renewals were considerably more common in the West than in the East. See U.S. Bureau of the Census (1895a, Vol. 12, pp. 3–5). Subsequent references to this volume are cited simply as 1890 Census.

letter of the law by setting a low interest rate and charging a high commission or by misrepresenting the interest rate on the mortgage document.[10] In Kansas, where the legal maximum was 12 percent, it was standard in the early 1870's to lend at 15 percent and to label the extra 3 percent as commission. However, there were cases, in Iowa and elsewhere, of a lender such as the Aetna Life Insurance Company of Hartford which lost foreclosure actions in court because of its practice of splitting the commission with its agent. (See Bogue, pp. 38; 50.)

Dissatisfaction was undoubtedly reinforced by visible attempts on the part of mortgage companies to reach collusive agreements to raise lending rates. Significantly, these attempts almost always quickly ended in failure. For example, in the autumn of 1880, the mortgage companies in Kansas arrived at an informal agreement under which their agents were instructed to combine and charge a common interest rate. In 1881, articles of agreement of the Kansas Loan Agents Association were drawn up and signed. Under their provisions, any lender who dropped his rate below that of the Association would be undercut by the other members until he complied. However, the Association proved incapable of eliminating undercutting and soon disappeared from the scene. Later in the decade, there was talk in Texas of combination to halt the decline in interest rates, but there too lenders proved unable to cooperate in the necessary action. (See Bogue, pp. 121; 160; passim.)

Risk was the mortgage lender's explanation for the high interest rates prevailing in the West. Many companies gave standard instructions to their agents in the field for appraising the riskiness of a mortgage loan. Agents were instructed to consider the characteristics of both the farmer and his land. Personal characteristics to consider included the farmer's marital status, the number and age of his children, whether he was of a nationality with a good reputation for farming, number of years he had been farming, and his record for meeting obligations. (See Kingman Robins, 1919, p. 75.) When evaluating the land, agents were instructed to consider soil type, elevation, topography, transportation facilities, and utilities. Climate was an important consideration in the semiarid regions of the West and Southwest. Soil type played an important role in determining the interest rates in other parts of the country. Thus, the sandhills along the Arkansas were forbidden to the agents of some companies and required higher interest rates for the agents of others. The same was true of other areas of river-bottom land throughout the American South. In the Far West, the moneylender found the rapid turnover of settlers and the unpredictable administration of law to increase the risks of lending. (See Bogue, p. 114.) Risks were perceived as greatest where few improvements were made on the farm and where farmers concentrated on growing one crop. (See Bogue; Virgil Lee, 1930, p. 288.)

Risk for the farmer implied risk for the lender to the extent that it increased the incidence of foreclosure. Foreclosing was costly due to court costs, transaction costs associated with selling the land, and interest earnings foregone in the interim. (See Robert Skilton, 1943, p. 320.) If the value of the land had declined since the loan was made, not only these costs but, in addition, any capital losses would be borne by the lender. This could occur either because of declining agricultural prices or because the farm deteriorated when left idle for a season. The response to this problem was to lend only a portion of the land's purchase price. In the rare instance that the value of the land against which the loan had been made appreciated prior to default, it was unusual for the lender to realize capital gains. In this event, the remainder of the sales price after the loan was repaid and foreclosure expenses were met typically went to the borrower.

The costs of foreclosing could be particularly significant in the 1890's, when judges with Populist sympathies allegedly permitted farmers' attorneys to disrupt foreclosure proceedings. Even when the lender managed to foreclose, it could be difficult to get full value for the land. To avoid problems of moral hazard, until the twentieth century some

[10] See 1890 Census, p. 167; Bogue, p. 28.

TABLE 2—REAL ESTATE MORTGAGE FORECLOSURES AS A PERCENTAGE OF MORTGAGES IN FORCE IN ILLINOIS, MINNESOTA, AND NEW JERSEY, BY YEARS

States	1886	1888	1889	1890
Illinois (acres)	0.61[a]	0.93		
Illinois (lots)	0.60[a]	0.67		
Minnesota (acres)	1.38[b]			1.55[c]
New Jersey (acres)			0.68	
New Jersey (lots)		0.76	0.64	

Source: 1890 Census. On the distinction between acres and lots, see Section IV below.
[a]Foreclosures of 1880.
[b]Foreclosures of 1881.
[c]Foreclosures of 1891.

states refused to permit the lender to purchase the land at foreclosure sale.[11] According to Bogue, it was rare indeed to find a disinterested individual willing to bid enough to extinguish the debt. To circumvent the prohibition on buying the foreclosed farm, mortgages were sometimes made by granting a trust deed. Title to the land was vested in a trustee nominally independent of the mortgage lender. If the terms of the mortgage were broken, the land was sold by the trustee and could be purchased by the actual lender, who could then rent it to a tenant farmer until an appropriate buyer was found.

The available information on foreclosure rates, while scanty, is summarized in Table 2. On foreclosure costs, data are available for the 65 mortgage loan companies licensed in New York, Connecticut, and Massachusetts in 1893; in that year foreclosure expenses amounted to 0.6 per cent of the value of loans secured by real estate (see Frederiksen, p. 211). Additional data for the period from 1914 on provide further information on foreclosure losses. A comprehensive study of farm mortgage foreclosure and loss experience conducted for the years 1914–26 showed that losses on loans ranged from 1/2 to 2 percent and were highest in such western states as Montana and Arizona.[12]

II. Modeling the Risk Premium

The purpose of this section is to develop a model that can be used to adjust mortgage interest rates for the risk premium. I wish to isolate the impact of risk on the mortgage interest rate, to purge that rate of the risk premium, and then to test whether there remains a significant interregional differential in the cost of mortgage credit.

In the paper closest in spirit to this one, James attempts to distinguish the roles of market power and risk differentials in the variation of interest rates at the end of the nineteenth and beginning of the twentieth centuries, using state-level data on the average rate of return on national banks' loan and discount portfolios. James regresses this rate of return against the variance of the loss rate on loans over the preceding five years (his measure of risk) and a number of other variables. A similar approach is adopted here. I wish to regress the mortgage interest rate against a measure of risk and other variables which may enter into the interest rate's determination.

The first step in modeling this problem is to define my concept of risk and to identify its sources. Following Knight, I take risk to be a situation in which alternative outcomes

[11]There is some dispute in the literature concerning the prevelance of such restrictions. Lee (p. 154) asserts that most states maintained such provisions into the twentieth century, while Skilton (p. 319) suggests that these restrictions had largely been eliminated by the 1830's. The difference may lay in the emphasis placed on foreclosure by court action: restrictions on lender purchases when land was foreclosed by court action declined much more rapidly than when land was foreclosed through other means.

[12]See F. F. Hill (1932) and Comptroller of the Currency (various issues). On the incidence of foreclosures during the interwar years, see Lee Alston (1983).

exist with known probabilities. The risks of mortgage lending arose from the probabilities of receiving alternative rates of return. In practice, a mortgage lender received a rate of return which differed from the mortgage interest rate when the borrower was in arrears. If the farmer simply made late payment, the lender might suffer a loss equal to foregone interest that would have been earned on the balance of the interim. However, mortgage instruments typically included provisions for the payment of penalties in such instances. Therefore, the risk to the lender arose largely from the possibility that the borrower might remain in arrears, forcing the lender to foreclose. The lender then ran the risk that the rate of return on the foreclosed loan would fall below the stated interest rate because sales proceeds proved insufficient to cover the principal, interest, legal fees, and transactions costs of reselling the land.

One way to proceed is to assume that the risk of lending is a function of the risk of farming and to regress the mortgage interest rate on one or more of the proxies for agricultural risk that have appeared previously in the literature. For example, we might adopt the coefficient of variation of wheat or cotton prices, the ratio of the variances of the transitory and permanent components of prices as calculated by Thomas Cooley and Stephen DeCanio (1977), or the variate difference measure of price uncertainty employed by McGuire (1980, 1981)

Each of these proxies is imperfect, and we can have little confidence that any of them adequately captures the risk of farming, much less the risk of mortgage lending. The variability, either total or unpredictable, of income from growing wheat or cotton is an inadequate measure of agricultural risk, since farmers could grow a diversified portfolio of crops, the returns on which would be imperfectly correlated, and thereby reduce the variability of returns.[13] Even were it feasible to construct an optimal portfolio of all the crops which might be grown, it would still have been possible for the grower to reduce the remaining risk by holding other assets (such as nonfarm real estate or securities) whose returns might be imperfectly correlated with those of farming.[14] Thus, the approach of constructing synthetic measures of risk is subject to important limitations.

An alternative is to build on the fact that the market prices risk. An acre of farmland is an asset, and its price contains information about the market's assessment of risk and expected returns to farming. I will need to provide a link from the risks of farming to the risks of mortgage lending.

Assuming that the risks of mortgage lending were a function of the probability of foreclosure, we can write the mortgage interest rate as

$$I = \rho + b_1 F + b_2 X, \tag{1}$$

where I is the mortgage interest rate and ρ is the risk-free rate of return on Treasury bonds or correspondent balances in New York.[15] The variable F is the risk of foreclosure, while X is a vector of other variables (to be introduced below) that enter into the mortgage rate's determination. Parameters b_1, b_2 and all the parameters that follow are defined to be positive. Consider a mortgage loan of given size. Assume that foreclosure takes place when income Y falls below the threshold level Y_F, and that this threshold level is less than the expected value of income $\bar{Y}$. Income Y should be understood to connote income over the span of time relevant to the foreclosure decision rather than, say, annual income.[16] For convenience, as-

[13] The extent to which farmers diversified their crops in response to economic incentives is the subject of considerable literature. On the Wheat Belt see DeCanio (1980), and on the Cotton South see Gavin Wright (1978).

[14] For examples of optimal portfolio calculations, see DeCanio and McGuire (1980).

[15] On the alternative assets available to lenders, see my earlier paper (1984).

[16] The assumption is that the decision to foreclose is made if the farmer is in default after a certain number of years. One can think of other plausible models of the foreclosure decision, such as, for instance, that foreclosure takes place when the farmer has been in arrears a certain number of times over the life of the mortgage or for a certain number of consecutive years. However,

sume Y to be normally distributed. Then the probability of foreclosure is a declining function of the expected value of income and an increasing function of its variance σ_Y^2:

$$(2) \qquad F = -c_1\overline{Y} + c_2\sigma_Y^2.$$

The argument here is that the price of land contains information about these variables. It is assumed that the price of an acre of farmland P depends positively on the expected value of income from cultivation and negatively on both that income's variance and the cost of mortgage credit:

$$(3) \qquad P = d_0\overline{Y} - d_1\sigma_Y^2 - d_2 I.$$

This relationship may be interpreted as follows. Risk-averse farmers will pay more for an acre of farmland the larger the returns from cultivation and the smaller the riskiness of those returns. If utility is quadratic, the utility of returns (and the valuation of land) will depend only on mean and variance. The term $d_0\overline{Y}$ is simply the present discounted value of expected income from cultivation; with an infinite horizon, d_0 is the reciprocal of the relevant discount factor. The term $d_1\sigma_Y^2$ is the adjustment for risk.[17] These variables can be measured in the same units as in (1) without loss of generality, assuming the utility function's additive separability across periods.

The final term in equation (3), $d_2 I$, permits the level of mortgage interest rates to be capitalized into the price of land. If farmers were concerned with the net returns from cultivation and if interest payments on mortgage debt were a significant component of costs, then farmers would have been willing to pay less for land where prevailing interest rates were higher. This could have been true even if the land was owned outright, so long as there was a positive probability that the land would have to be resold to a buyer requiring a mortgage. However, except where mortgages were assumable, the average interest rate prevailing in the local mortgage market on loans on land with the relevant characteristics, as opposed to the interest rate actually paid by the current borrower, was likely to be capitalized into the price. Thus, we are more likely to observe this capitalization effect in state level data, given the wide variation in mortgage interest rates in different regions, than in my data for Wisconsin, in which there is less variation in the characteristics of land and where the variation in interest rates would have been associated to a greater extent with farmers' personal characteristics. In each case, I will want to test for this effect explicitly.

Substituting (2) and (3) into (1) yields

$$(4) \qquad I = \phi\rho + \phi\alpha\overline{Y} - \phi\gamma P + \phi b_2 X,$$

where $\phi = \dfrac{d_1}{d_1 + b_1 c_2 d_2} > 0,$

$$\alpha = b_1\left[\frac{c_2 d_0}{d_1} - c_1\right] \gtrless 0,$$

and $\gamma = b_1(c_2/d_1) > 0.$

Thus, the coefficient on land prices should enter equation (4) with a negative sign, while the coefficients on yields may be either positive or negative. From (2), higher yields reduce foreclosure risk holding variability constant, with a negative effect on interest rates; but from (3), higher yields imply higher variability holding land price constant, which

implementing such rules in an empirical framework would require information on the serial correlation of the yields on optimal portfolios, which would be extremely difficult to estimate in practice.

[17] In principle, one might argue that the price of land should be an increasing function of the variance of the income stream, given my assumption about foreclosure. Assume that the worst the farmer could do when income fell below Y_F was to be foreclosed upon. Then farmers would not have suffered more from realizations of income significantly below Y_F than from draws just equaling Y_F, while in contrast they would have benefited significantly from draws well into the upper tail of the distribution. Thus, farmers would have preferred lands where income was highly variable. However, this argument holds only in the absence of significant costs of relocation. Its validity hinges upon the unrealistic assumption that a farmer experiencing foreclosure could, at low cost, relocate and obtain a new mortgage. This assumption's lack of realism and contemporary farmers' constant complaints regarding uncertainty led me instead to adopt the formulation in the text.

increases foreclosure risk and puts upward pressure on interest rates. Note that in equation (4), $\alpha\bar{Y}-\gamma P-(1/\phi-1)I$ is a measure of the risk premium. A regression of I on a constant term, $\bar{Y}$, P, and X would provide us with estimates of α and γ if we knew the value of ϕ.[18] However, the constant term equals $\phi\rho$. My strategy will be to utilize information on the rates of return on safe assets to impose values for the risk-free rate ρ, and to use these parameters in conjunction with the estimated constant terms to derive values for α and γ. Ultimately, none of my conclusions will hinge upon the value chosen for ρ.

Obviously, the model is based on a number of restrictive assumptions. A first set of assumptions concerns the diversification of risk. I do permit farmers to hedge certain types of risks by growing a diversified portfolio of crops. To the extent that this strategy was feasible, the reduction in risk would be reflected in land values. However, I do not explicitly model the process whereby mortgage companies could have hedged foreclosure risk when setting lending rates. In theory, lenders could have considered whether a particular farmer was more or less likely to fail during periods when other farmers to whom they had loans outstanding were also likely to fail, and they could have charged a higher interest rate the higher the correlation. In fact, the evidence on loan administration presented above suggests that agents followed mechanical rules and evaluated each loan individually. These rules of conduct may have incorporated the head office's assessment of the covariance of risks and been an attempt to guide agents to act as if they took such covariances into account. However, the model only incorporates hedging behavior by banks and mortgage companies to the extent that the covariation of risks was reflected in land values.

Second, I have not specified foreclosure risk as a function of the size of the mortgage debt. In fact, the ratio of mortgage debt in force per mortgaged acre to average value per acre is remarkably stable across regions.[19] If the mortgage debt is a constant proportion of the land's value, then its omission will cause this effect to be incorporated into the coefficient on land values, reducing multicolinearity but leaving the results otherwise unaffected.[20] Of course, this is a testable proposition, and I return to it below.

III. Evidence from Wisconsin

To test for the equality of risk-adjusted mortgage interest rates across regions of the country, there is no choice ultimately but to rely on state-level data drawn from the 1890 Census. Since economic and climatic conditions varied widely within states, some readers may harbor the suspicion that an accurate measure of the risk premium cannot be isolated at such high levels of aggregation. Ideally, one would wish to examine the determinants of mortgage interest rates using observations from a representative sample of individual mortgage loans in various parts of the country. Unfortunately, no such sample appears to be available. However, a number of state and local agencies conducted surveys in the last decades of the nineteenth century in which mortgage interest rates were gathered.[21] Before turning to the aggregate data, let us utilize in this section one such survey, that of Wisconsin farmers in 1895.

[18] Were there no capitalization of interest rates, then $d_2=0$ and $\phi=1$, so no adjustment would be required. In this case, the system (1)–(3) is recursive, and (4) could be estimated by ordinary least squares. If $d_2>0$, then the system is simultaneous, and estimation of (4) will require instruments.

[19] For the 22 states for which the Census Bureau made this comparison, the ratio was 38.27 percent in the South Atlantic, 37.62 percent in the North Central, 37.76 percent in the South Central, and 40.56 percent in the West. The simple correlation across states of land prices and mortgage incumbrance is 0.96.

[20] Denote debt per mortgaged acre D. Then (2) becomes $F=-c_1\bar{Y}+c_2\sigma_Y^2+c_3D$, where $D=e_1P$. Substituting as in the text, (4) becomes

$$I=\phi\rho+\phi\alpha\bar{Y}-\phi(\gamma+b_1c_3e_1)P+\phi b_2X.$$

[21] Most of the surveys conducted prior to 1890 are mentioned in 1890 Census, Vol. 14, pp. 3–5.

Agrarian radicals were active in Wisconsin from the close of the Civil War through the turn of the century. Wisconsin farmers were among the nation's best organized during the Granger period (1867–74), when they engaged in largely unsuccessful attempts to regulate railroad rates and grain elevator practices.[22] Wisconsin's agricultural interests remained exceptionally active during the Greenback period, which extended from the 1873 financial crisis into the early 1880's.[23] The Greenbackers were primarily concerned with monetary issues and lobbied for higher prices in general and higher agricultural prices in particular. Compared to the Greenbackers, members of the Alliance Movement of the 1880's were less active in Wisconsin. The center of protest and the effort to form agricultural cooperatives designed to bypass middlemen were located to Wisconsin's north and west, in Minnesota, the Dakotas, Kansas, Nebraska, Iowa, and Missouri. Nationally, the peak of agrarian unrest was reached in the 1890's with the rise of the Populist Party. The Populists were concerned with the monetary system, railroad regulation, and the level of mortgage interest rates; their concern with the latter was reflected in the sequence of inquiries into the level of interest rates conducted in the 1890s. While Wisconsin was not one of the principal strongholds of Populist sentiment, it was nonetheless one of the midwestern states in which the Populists found support.[24]

Wisconsin had a long history of usury laws stretching back to 1839; throughout this period Wisconsin farmers lobbied for their retention. The interest ceiling, which had been lowered from 10 to 7 percent in 1863, was restored to 10 percent in 1866 (1890 Census, pp. 172–83). The penalty in the event of violation was loss of interest and forfeiture of triple the amount of the illegal interest. These laws were maintained through the 1890's, although in 1876 building and loan associations were exempted from their provisions.

In 1895 the Wisconsin State Bureau of Labor, Census and Industrial Statistics conducted an inquiry into the condition of the state's farmers. The Bureau sent out some 5,000 inquiries to farmers throughout the state, of which approximately 10 percent were returned. While the Bureau noted that the investigation "did not meet with the encouragement expected," it took solace in the fact that the farmers' statements "were evidently prepared with great care, and together with the opinions given in answer to questions of a more general nature... represent quite fully what may be considered the average social and economic condition of the farmers in Wisconsin, as well as their individual views upon social and economic topics" (p. 2). The Bureau's questions included a number of items of personal information, such as age, marital status, number of children, nativity, and number of years a farmer; a number of items of information about the farm, such as number of acres, value per acre including improvements and yield per acre of various crops; and a number of items of farming practice, including fertilizer use, men and women employed, wages paid, fire insurance, and whether the farmer had saved or run into debt in the previous year. No information on the amount of the mortgage incumbrance was gathered. The Bureau's *Seventh Biennial Report* prints the 549 returns: 194 farmers, or 35 percent of the total, provided an answer to the question, "If farm is mortgaged what rate of interest do you pay?" In no instance did the reported mortgage interest rate exceed the usury ceiling. Of those 194 farmers, 180 provided all of the information required for regression analysis.

Despite the Bureau of Labor Statistics' assurance that the survey responses accurately reflect the average condition of Wisconsin farmers, the relatively low response rate raises the possibility of sample selection bias. An informal check is to compare the survey figures with aggregates from the 1890 Census. Of the survey respondants, 35.33 percent answered the question, "If farm

[22] See Buck (pp. 211–89); Vernon Carstensen (1974, pp. 17–68); Shannon (pp. 300–26).

[23] Irwin Unger (1964, pp. 345–407); Lawrence Goodwyn (1976, pp. 16–24).

[24] Hicks (pp. 225–65); Richard Hofstadter (1955, pp. 94–109); Robert McMath (1975, pp. 106–09).

is mortgaged what rate of interest do you pay?" In the four Wisconsin counties for which the Census Office calculated mortgaged acres as a percentage of all (taxed) acres, the figure was 36.90 per cent.[25] Comparing the census figure for 1889 on the average interest rate of mortgages on acres with the average reported by the survey respondants in 1895, the 1889 figure of 6.75 percent is broadly consistent with the 1895 figure of 6.3 percent, considering the downward trend in rates.[26]

Estimation of (4) requires a measure of expected income from cultivation. I use each farmer's estimates of average annual yield per acre of wheat, oats, barley, corn, potatoes, buckwheat, tobacco, and hay—all of the crops for which the Wisconsin State Bureau of Labor reported the relevant data —along with the percentage of the land cultivated. Employing physical yields rather than their value entails an assumption about prices. Let us assume that any deviation of the price received by a farmer from that year's average price is random, on the grounds that markets were well integrated.[27] Therefore, the coefficients on crop yields in the tables presented below should be interpreted as the product of average crop prices and the structural parameters in equation (4).[28]

To test for the endogeneity of land prices due to the capitalization of interest rates, I calculated Jerry Hausman's m-statistic, finding no evidence of endogeneity.[29] Therefore, the equations in Table 3 are estimated by ordinary least squares. Table 3 reports a number of regressions of the mortgage interest rate on the value per acre of farmland including improvements, measures of farm output, and farmers' personal characteristics.

TABLE 3—REGRESSION RESULTS FOR WISCONSIN FARMERS, 1895

	(3.1)	(3.2)	(3.3)
Constant	7.39	6.96	7.18
	(31.61)	(21.86)	(20.54)
Land Value	−0.012	−0.014	−0.013
	(3.62)	(3.98)	(3.84)
Wheat	−0.003	−0.002	−0.003
	(0.47)	(0.33)	(0.44)
Oats	−0.009	−0.009	−0.009
	(1.53)	(1.54)	(1.48)
Barley	−0.006	−0.006	−0.006
	(1.27)	(1.31)	(1.24)
Corn	0.005	0.006	0.006
	(1.39)	(1.73)	(1.71)
Potatoes	0.001	−0.001	−0.001
	(0.31)	(0.06)	(0.22)
Buckwheat	0.020	0.016	0.016
	(2.20)	(1.81)	(1.80)
Tobacco	0.001	0.001	0.001
	(0.49)	(0.72)	(0.80)
Hay	−0.211	−0.190	−0.180
	(2.18)	(2.00)	(1.90)
Save		−0.280	−0.241
		(2.12)	(1.80)
Debt		0.314	0.289
		(2.07)	(1.88)
Experience		0.007	0.007
		(1.48)	(1.52)
Percent cultivated		0.417	0.466
		(1.27)	(1.41)
Fertilizer			0.189
			(0.92)
Fire			−0.357
			(1.56)
R^2	.217	.284	.297

Notes: The dependent variable is mortgage interest rate; t-statistics are shown in parentheses; number of observations is 180.

[25] This is an arithmetic average of the percentages for Brown, Douglas, St. Croix, and Waukesha counties; see 1890 Census, p. 125.

[26] See 1890 Census, p. 249. In addition, part of the discrepancy may be attributable to the way the Census Office adjusted interest rates for commission charges. See Section IV below for details.

[27] Evidence on market integration is discussed by Philip Coelho and James Shepherd (1974).

[28] To the extent that the period of time relevant to the foreclosure decisions was longer than a year, an additional adjustment is required. For example, if a two-year period was relevant to the foreclosure decision, average yields per two-year period are the preferred regressors. But since annual yields are one-half the relevant regressor, the coefficient estimates must simply be further divided by two to recover the structural parameters.

[29] See Hausman (1978). I employed in various combinations the other variables included in the Wisconsin survey, such as marital status, family size, men and women employed, wages paid, and membership in labor or beneficiary organizations. In no instance did I reject the null hypothesis of exogeneity of P at the 90 percent level of confidence.

In equation (3.1), the coefficient on value per acre has a negative sign as the underlying model would predict, and the yield variables are significant as a group at the 95 percent level of confidence. Equations (3.2) and (3.3) add other characteristics of the farm and its operator (X variables). Farmers were asked separate questions about whether they managed to save during the previous year or ran into debt; none in this sample answered both questions affirmatively, but some answered no to both which can be taken to imply that these farmers broke even. In equation (3.2), both the saving and debt variables are significant, suggesting that farmers who had saved obtained favorable interest rates, while farmers who had run into debt paid higher ones. Years a farmer enters with a surprising positive sign, but the coefficient is insignificantly different from zero. The zero coefficient may indicate that this variable is capturing both the impact of time spent farming as a signal of reputability and the countervailing fact that farmers with greater tenure may have acquired their mortgages earlier, during years when higher interest rates prevailed. The coefficient on the percentage of the acreage cultivated also is insignificantly different from zero, which may be explicable on the grounds that the decision whether to cultivate was a function of characteristics of the land which were incorporated into the value per acre. In equation (3.3), dummy variables for fertilizer use and fire insurance are added; the coefficient on the former is insignificant, while the coefficient on the latter borders on significance and has a plausible negative sign.

IV. Evidence from the 1890 Census

Results of analyzing survey evidence from Wisconsin are in broad accord with the predictions of my model of the determinants of mortgage interest rates. In this section, I reestimate the model on state-level data drawn from the 1890 Census to provide a basis for comparing risk-adjusted interest rates in different parts of the country.

Under the provisions of an act of Congress instructing the Superintendent of the Census to collect "statistics of, and relating to, the recorded indebtedness of private corporations and individuals," a massive investigation of real estate mortgages was conducted for the *Eleventh Census* (see p. 3). The demand for information regarding mortgage debt was said in the Census Report to have grown out of "widespread discontent among farmers and workingmen and the discussions of persons interested in social science" (p. 3). The only survey of real estate mortgages concluded before the Census Office began its work was conducted by the Illinois State Bureau of Labor Statistics for 1870, 1880, and 1887 (published 1888). The Census Office limited its study to real estate mortgages, excluding crop liens, chattel mortgages, and judgments by process of law. Although a special inquiry was conducted by mail for 102 selected countries, the bulk of the information was gathered by special agents from public mortgage records in county real estate record offices.

The Census Office's investigation was unusually comprehensive. An attempt was made to gather information for every outstanding real estate mortgage. The government's special agents reported encountering a variety of problems. Some agents had to read records where handwriting was almost illegible or the ink was barely visible. Records were sometimes poorly organized and maintained. Records written in French were found in Louisiana, in German in Illinois, and in Spanish in New Mexico. In New Mexico, debt was sometimes denominated in number of sheep, cattle, or horses, and the money value of the stock had to be calculated. In recently settled parts of the West where counties sometimes had two public recorders until the courts established the title of one, mortgages could be recorded in both offices, and agents had to make a special effort to avoid counting duplicates. Finally, there was the problem of fraudulent mortgages, typically given to friends or business acquaintances who were not creditors in order to protect the mortgagors' real estate from the claims of acutal creditors. Fraudulent mortgages were identified when, for example, the agent discovered that the mortgage covered real estate that could not be located, or was worth only a fraction of the purported debt. Such

mortgages were eliminated from the agents' records.

Real estate was classified as acres or lots, and mortgage interest rates were tabulated separately for each category. Lots included real estate situated in cities and smaller towns and in their vicinity if withdrawn from agriculture and subdivided or intended for urban use. Acres included all other real estate, chiefly agricultural but including also pasturage, woodland, mining land, and certain suburban tracts not subdivided into lots. While the Census Office recognized the desirability of distinguishing agricultural and nonagricultural land, the task of classification proved impossible given the available resources. In this section I utilize data on acres, which therefore covers chiefly but not solely agricultural land. Information on the percentage of land cultivated can provide an indication of the share of that land devoted to pasturage and woodland.

While most of the information gathered by the Census Office was tabulated at the county level, statistics on mortgage interest rates are reported only for states. Again, this information appears to have been gathered with considerable care. However, the investigators encountered a variety of problems when attempting to record the mortgage interest rate. Sometimes the interest rate was reported not on the mortgage instrument but on a promissory note to which the mortgage referred. In the South and West mortgages were found which stated that the debt was without interest or mentioned a lower interest rate than the true one, and the principal was artificially inflated to make up the difference. When such "stuffed mortgages" were identified, the principal was reduced and the interest payment was increased accordingly. Between the Mississippi and the Rockies, commission mortgages were common. The Census presents an example of a mortgage loan at 7 percent interest to which was attached a second, smaller mortgage with no interest, which represented a 3 percent commission. Since commissions were normally paid out of the interest charge on the mortgage, the Census Office required that the rate of interest should be inclusive of commission for the sake of uniformity. Agents were instructed to record all such commission mortgages as interest rather than principle.

The Census Office's 1889 figures for interest rates and mortgage incumbrance are linked to other information for the same year drawn from the *Eleventh Census* Report of Agriculture (1895b). I employ data on the value per acre of farmland and improvements, percentages of farmland improved and cultivated, yields per acre of various crops, and expenditures on fertilizer. For yields per acre, I use the same crops as in Section III, but with the addition of cotton.[30] However, in place of farmers' estimates of typical yields, I use actual yields for calendar year 1889. It would be preferable, of course, to generate expected yields from time-series on actual yields and other information useful for predicting farm incomes, but time-series of the required length are not available on a national basis for a sufficient variety of crops. The use of actual yields will pose a problem only if 1889 was an atypical year in certain regions but not in others. If hay yields everywhere were a certain percentage above average in 1889, the hay coefficient would be underestimated by the same percentage but, since the product of the coefficient and the yield are unaffected and that product is used to calculate the risk premium, the estimate of the premium itself would be unchanged. Only if 1889 was an atypical year in a particular state or region but not in others will my use of actual yields be a source of bias. A reading of the literature fails to turn up exceptional instances of insect infestation or natural disaster in 1889. However, 1889 was part of a period of adjustment to declining average precipitation in the region stretching west from the central portions of Kansas, Nebraska, and the Dakotas. The period prior to 1886 had been one of abundant rainfall, but 1887 was the first of a stretch of dry years lasting for more than a decade. This would have posed a problem if farmers had

[30] This raises the question of how to handle states where no cotton was grown. Entering an average yield per acre of zero seems appropriate in terms of the theory, since the fact that the crop was not grown can be taken to indicate that, were it grown, yields would have been exceptionally low.

TABLE 4—REGRESSION RESULTS BY STATE, 1889

	(4.1)	(4.2)	(4.3)
Constant	3.648	3.467	3.201
	(1.56)	(1.43)	(1.03)
Land Value	−0.053	−0.065	−0.071
	(2.54)	(2.01)	(1.57)
Wheat	0.093	0.093	0.087
	(1.52)	(1.52)	(1.14)
Oats	0.035	0.032	0.036
	(0.81)	(0.74)	(0.70)
Barley	−0.109	−0.104	−0.100
	(2.15)	(1.97)	(1.67)
Corn	−0.061	−0.066	−0.065
	(2.28)	(2.27)	(2.17)
Potatoes	0.019	0.018	0.018
	(1.84)	(1.81)	(1.50)
Buckwheat	−0.037	−0.033	−0.033
	(1.22)	(1.02)	(1.00)
Rye	0.094	0.102	0.105
	(1.55)	(1.60)	(1.55)
Hay	1.644	1.729	1.781
	(2.73)	(2.71)	(2.46)
Cotton	2.413	2.309	2.314
	(1.27)	(1.19)	(1.14)
Tobacco	−0.001	−0.001	−0.001
	(2.54)	(2.38)	(2.24)
Percent Cultivated	0.027	0.026	0.028
	(1.36)	(1.25)	(1.09)
Percent Improved	2.258	2.710	2.926
	(1.35)	(1.41)	(1.25)
Incumbrance		0.014	0.011
		(0.51)	(0.33)
Fertilizer			0.452
			(0.16)
Standard Error of the Regression	0.885	0.906	0.938

Notes: The dependent variable is mortgage interest rate; t-statistics are shown in parentheses; number of observations is 47.

not revised their expectations in light of this new information about rainfall by 1889. In fact, Hicks (pp. 32–33) suggests that expectations about yields were transformed quite rapidly.[31]

The number of observations is 47, since Oklahoma lacks the necessary mortgage information and the District of Columbia is excluded on the grounds that the sample of agricultural land is small and unrepresentative.[32] I again tested the exogeneity of land values and mortgage debt by calculating Hausman's m-statistic, using the shares of acreage devoted to the various crops as instruments. These tests rejected the null hypothesis of exogeneity at the 95 percent level of confidence.[33] The regressions are therefore estimated by instrumental variables. Results are shown in Table 4. As in Table 3, equation (4.1) reports a regression of the mortgage interest rate on land value per acre including improvements, average crop yields per acre, and percentages of land cultivated and improved. Once again, value per acre has a negative sign as the theory predicts, while the yield variables are significant as a group. Equations (4.2) and (4.3) add the mortgage incumbrance per mortgaged acre and the value of purchased fertilizer per acre. The incumbrance is insignificant. The fertilizer variable, while insignificant, has the same positive sign as in Table 3. Percent cultivated has a positive sign but is insignificant, as in Table 3. Percent improved is also insignificant, consistent with the presumption that improvements should show up in value per acre.

V. Testing for Interregional Differentials

We are now in a position to test for the presence of a regional differential in mortgage interest rates. Consider first the unadjusted rates. Mortgage rates averaged 5.81 percent in the North Atlantic states of Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, and Pennsylvania compared to 8.17 percent in the rest of the country. The difference in means is statistically significant at the 99 percent level of confidence.[34] If we

[31] To the extent that mortgages were issued before 1887 on the basis of expected yields far different from those reflected in Census data for 1889, it would be desirable to exclude these early mortgages from my sample. Unfortunately, this is not permitted by the data.

[32] For instance, mortgage debt per acre in the District was 23 times the national average and $3\frac{1}{2}$ times that of the highest state.

[33] An explanation for the contrast with the result in Section III is suggested in Section II.

[34] The test statistic of 4.67 is distributed as a student's t with 45 degrees of freedom.

add to the North Atlantic Division the South Atlantic states of Delaware, Maryland, Virginia, West Virginia, and the Carolinas, then mortgage interest rates average 6.14 percent for the Atlantic states compared to 8.46 percent in the rest of the country. This difference is also significant.[35]

The next step is to use the coefficients in Table 4 to calculate the risk premium.[36] This requires a value for ρ. In 1889, U.S. 4 percent bonds were trading on average at 128 percent of par and hence yielding 3.1 percent.[37] This is a reasonable lower bound for ρ. Frederick Macauley's unadjusted index of American railroad bond yields averaged 4.4 percent in 1889 (1938, Appendix Table D). Since there was some risk of default on railroad bonds, this is certainly an overestimate of the risk-free rate, so I use 4.4 percent as an upper bound for ρ. By its definition in Section II above, ϕ lies between zero and one. Therefore I calculate the risk-adjusted mortgage rate using values for $1/\phi$ of 1.0, 1.1, and 1.2 in conjunction with equation (4.1) in Table 4.[38] This yields estimates for ρ of 3.65, 4.01, and 4.38. The risk-adjusted mortgage interest rates are shown in Table 5 along with the unadjusted rates, while the risk premia are shown in Table 6. As demonstrated below, none of the conclusions of this section are altered by assuming different values for ρ.

The figures in Table 5 can be used to test the significance of regional differences in risk-adjusted mortgage rates. The adjusted rates can simply be regressed on a constant term and a dummy variable for the North Atlantic states. However, since a majority of states had interest rate ceilings in effect and ceilings set at certain levels typically were concentrated in certain regions, we will want to control for the presence of interest rate ceilings when testing for a regional effect. There exists the possibility that low mortgage rates in some regions were associated with binding interest rate ceilings. The eleven states with 6 percent ceilings were all located in North Atlantic and South Atlantic regions, while the 12 and 18 percent ceilings were applied in western states (North Dakota, South Dakota, Texas, and Idaho). As described above, it was sometimes possible to circumvent these laws through the use of commission charges and other forms of subterfuge, so it is by no means certain that the usury laws had a significant impact on interest rates. Since the Census Office's special agents were instructed to add commissions and other hidden charges to the interest rate, to the extent that evasion took this form and could be discerned by the agents, we are provided with an accurate measure of the interest rate. The Census found that 1.22 percent of all mortgages, amounting to 0.67 percent of the value of the debt, were in violation of the usury laws.[39] (See Table 7.)

It is not clear how this number should be interpreted. If 1.22 percent is seen as a large number, this may imply that the usury laws were freely violated and had no impact on interest rates. Conversely, if 1.22 percent is seen as a small number, this could mean that the interest rate ceilings were set at such high levels that they were of little relevance, in which case they again would have had no impact on interest rates. Alternatively, if 1.22 percent is seen as a small number, this could

[35]The t value is 5.93.

[36]Clearly, this procedure hinges upon the assumption that the parameters estimated in the previous section capture the impact of risk on the probability of foreclosure and through that channel on the cost of mortgage credit. It is conceivable, as in any econometric study, that the estimated coefficients are picking up the effects of other factors correlated with land prices or yields but omitted from the equation. For example, it is conceivable that lenders discriminated against farmers of land worth relatively little or growers of certain crops for reasons independent of risk. In principle, this hypothesis could be tested by including in equation (4) dummy variables for the crops or lands in question; in practice, it is hard to think of independent evidence that could be used to guide the construction of such variables.

[37]U.S. Comptroller of the Currency (1891, Vol. 1, p. 50).

[38]The conclusions to this section are unaltered if I use other equations in Table 4.

[39]See 1890 Census, p. 174. By far the highest incidence of usurious mortgages was in Alabama, Georgia, and Kentucky. The western states and territories with the largest share of usurious mortgages were New Mexico and Oregon.

TABLE 5—ACTUAL AND RISK-ADJUSTED INTEREST RATES

	R	$ALT1$ ($1/\phi = 1$, $\rho = 3.65$)	$ALT2$ ($1/\phi = 1.1$, $\rho = 4.01$)	$ALT3$ ($1/\phi = 1.2$, $\rho = 4.38$)
Maine	6.240	2.161	2.377	2.594
New Hampshire	5.980	4.050	4.455	4.860
Vermont	5.920	3.435	3.778	4.122
Massachusetts	5.600	4.270	4.698	5.125
Rhode Island	5.770	3.896	4.286	4.675
Connecticut	5.720	3.843	4.228	4.612
New York	5.710	3.552	3.907	4.262
New Jersey	5.660	3.751	4.126	4.501
Pennsylvania	5.650	4.359	4.795	5.231
Delaware	5.770	3.093	3.403	3.712
Maryland	5.900	3.080	3.388	3.696
Virginia	5.930	2.212	2.433	2.654
West Virginia	5.990	2.647	2.911	3.176
North Carolina	7.750	3.274	3.602	3.929
South Carolina	8.560	4.903	5.394	5.884
Georgia	8.090	3.696	4.065	4.435
Florida	9.120	4.131	4.544	4.957
Ohio	6.590	4.082	4.490	4.898
Indiana	6.670	3.825	4.208	4.590
Illinois	6.700	3.635	3.999	4.363
Michigan	7.020	3.460	3.806	4.152
Wisconsin	6.750	3.487	3.836	4.185
Minnesota	7.960	2.530	2.783	3.036
Iowa	7.410	3.467	3.814	4.160
Missouri	8.040	3.979	4.377	4.775
North Dakota	9.280	4.073	4.481	4.888
South Dakota	9.040	4.110	4.521	4.932
Nebraska	8.190	3.948	4.343	4.737
Kansas	8.380	3.138	3.452	3.765
Kentucky	6.310	3.566	3.923	4.280
Tennessee	5.990	2.511	2.762	3.013
Alabama	8.340	3.218	3.540	3.862
Mississippi	9.540	4.323	4.756	5.188
Louisiana	8.010	3.243	3.567	3.891
Texas	9.130	3.657	4.023	4.389
Arkansas	9.030	4.005	4.406	4.806
Montana	10.790	5.438	5.982	6.526
Wyoming	9.540	3.989	4.388	4.787
Colorado	8.530	3.291	3.620	3.949
New Mexico	7.900	2.411	2.652	2.893
Arizona	11.350	4.316	4.748	5.180
Utah	9.590	3.824	4.207	4.589
Nevada	10.840	3.249	3.574	3.899
Idaho	10.280	4.466	4.912	5.359
Washington	8.000	2.185	2.404	2.622
Oregon	9.280	4.152	4.567	4.982
California	9.030	5.508	6.059	6.610

also mean that interest rate ceilings were set at low levels and were effectively enforced, in which case they would have had a significant impact on interest rates. The regression results reported below shed some light on their significance.

Separate dummy variables, defined for states with interest rate ceilings set at 6, 8, 10, 12, and 18 percent, are denoted *MAX6*, *MAX8*, *MAX10*, *MAX12*, and *MAX18*. In addition, I define *EAST* as a dummy variable taking on a value of unity for the nine

TABLE 6—ESTIMATES OF THE RISK PREMIA

	*PREM*1 ($1/\phi=1$, $\rho=3.65$)	*PREM*2 ($1/\phi=1.1$, $\rho=4.01$)	*PREM*3 ($1/\phi=1.2$, $\rho=4.38$)
Maine	4.078	3.862	3.645
New Hampshire	1.929	1.524	1.119
Vermont	2.484	2.141	1.797
Massachusetts	1.329	0.901	0.474
Rhode Island	1.873	1.483	1.094
Connecticut	1.876	1.491	1.107
New York	2.157	1.802	1.447
New Jersey	1.908	1.533	1.158
Pennsylvania	1.290	0.854	0.418
Delaware	2.676	2.366	2.057
Maryland	2.819	2.511	2.203
Virginia	3.718	3.496	3.275
West Virginia	3.342	3.078	2.813
North Carolina	4.475	4.147	3.820
South Carolina	3.656	3.165	2.675
Georgia	4.393	4.024	3.654
Florida	4.988	4.575	4.162
Ohio	2.507	2.099	1.691
Indiana	2.841	2.461	2.079
Illinois	3.064	2.700	2.336
Michigan	3.559	3.213	2.867
Wisconsin	3.262	2.913	2.564
Minnesota	5.429	5.176	4.923
Iowa	3.942	3.595	3.249
Missouri	4.060	3.662	3.264
North Dakota	5.206	4.798	4.391
South Dakota	4.929	4.518	4.107
Nebraska	4.241	3.846	3.452
Kansas	5.241	4.927	4.614
Kentucky	3.478	3.227	2.976
Tennessee	2.743	2.386	2.029
Alabama	5.121	4.799	4.477
Mississippi	5.216	4.783	4.351
Louisiana	4.766	4.442	4.118
Texas	5.472	5.106	4.740
Arkansas	5.024	4.623	4.223
Montana	5.351	4.807	4.263
Wyoming	5.550	5.151	4.752
Colorado	5.238	4.909	4.580
New Mexico	5.488	5.247	5.006
Arizona	7.033	6.601	6.169
Utah	5.765	5.382	5.000
Nevada	7.590	7.265	6.940
Idaho	5.813	5.367	4.920
Washington	5.814	5.595	5.377
Oregon	5.127	4.712	4.297
California	3.521	2.970	2.419

North Atlantic states. Including dummy variables for both regions and states with interest rate ceilings minimizes the likelihood that one type of variable is picking up effects associated with the other. In addition, since the dependent variable has already been adjusted for the effects of variations in land values and yields, it is unlikely that other variables omitted from these equations but correlated with regional differences in land values will be picked up by the dummy variables' coefficients.

TABLE 7—NUMBER, VALUE, AND PERCENTAGE OF USURIOUS REAL ESTATE MORTGAGES, 1880–89

	Total Mortgages		Usurious Mortgages		Usurious as Percent of Total Mortgages	
	Number	Value	Number	Value	Number	Value
United States	9,517,747	$12,094,877,793	116,254	$80,748,780	1.22	0.67
All States and Territories						
North Atlantic	2,487,490	4,819,223,581	3,151	5,960,329	0.13	0.12
South Atlantic	620,400	606,558,768	26,189	16,238,862	4.22	2.68
North Central	5,003,184	4,767,404,337	20,092	13,115,667	0.40	0.28
South Central	848,294	888,263,139	63,404	42,258,647	7.47	4.76
Western	558,379	1,013,427,968	3,418	3,175,284	0.61	0.31
States and Territories Having Usury Laws						
North Atlantic	2,025,260	4,049,323,033	3,151	5,960,329	0.16	0.15
South Atlantic	578,443	569,209,690	26,189	16,238,862	4.53	2.85
North Central	5,003,184	4,767,404,337	20,092	13,115,667	0.40	0.28
South Central	785,693	763,687,959	63,404	42,258,647	8.07	5.53
Western	70,450	100,103,698	3,418	3,175,284	4.85	3.17

Source: 1890 Census.

TABLE 8—TESTING FOR A REGIONAL DIFFERENTIAL IN RISK-ADJUSTED MORTGAGE INTEREST RATES

	(8.1)	(8.2)	(8.3)
Constant	3.635	3.811	3.729
	(29.58)	(18.92)	(16.83)
East	0.067		0.286
	(0.23)		(0.89)
MAX6		−0.515	−0.563
		(1.69)	(1.82)
MAX8		−0.189	−0.107
		(0.51)	(0.28)
MAX10		−0.047	0.034
		(0.16)	(0.11)
MAX12		−0.25	−0.166
		(0.58)	(0.38)
MAX18		0.66	0.74
		(0.84)	(0.94)
R^2	.01	.10	.12

Notes: The dependent variable is the risk-adjusted interest rate from Table 5, col. 2; *t*-statistics are shown in parentheses; number of observations is 47.

Tests for the presence of a regional differential and for the effects of usury laws are presented in Table 6. Regressions are run entering the regional and interest ceiling variables both separately and together. Only results using the first set of risk-adjusted mortgage rates in Table 5, calculated for $1/\phi=1$, are presented in the table. Since the risk-adjusted rates in the last two columns of Table 5 are a linear transformation of those considered in the first column, the reader can derive the analogous results for the other values of ϕ by multiplying each coefficient and standard error in Table 8 by 1.1 or 1.2.

The three regressions tell a consistent story. With the risk premium removed from the mortgage rates, there remains no statistically significant regional differential. However, there is some evidence suggesting that interest rate ceilings when set at low levels depressed mortgage interest rates. Either three or four of the five interest ceiling variables are negative in sign, but only the 6 percent ceiling variable differs significantly from zero at the 90 percent level or better.[40] The coefficient on this variable suggests that, for $\phi=1$, the presence of a 6 percent ceiling lowered interest rates by approximately half a per-

[40]The danger with dummy variables is that they may be picking up characteristics of states other than the effects of their usury laws. This is clearly the case with *MAX18*, since there is only one state (Idaho) with an 18 percent interest rate ceiling. However, each of the other dummy variables covers a number of states drawn from two or more regions. The highest regional concentration is for the 6 percent ceilings, but they apply in only five out of nine North Atlantic states and four out of nine South Atlantic states.

centage point.[41] In any case, once mortgage interest rates have been adjusted for risk and for the effects of statutory ceilings, there remains no evidence of a significant interregional differential.

VI. Conclusions and Implications

This paper has attempted to account for regional differences in mortgage interest rates during the Populist era. Two conflicting hypotheses dominate the literature on this subject. The first is based on the farmers' complaint that imperfectly competitive market conditions in the American West permitted mortgage lenders to collude in raising interest rates. The second is based on evidence of competition and on the observation that interest rate differentials could in theory reflect the existence of risk premia that varied by region of the country. The strategy of this paper has been to adjust mortgage interest rates for such risk premia and then to test for the presence of residual variation across regions. While differences in risk have been found to exercise a significant influence over the level of interest rates, adjusting those rates for the impact of risk does not eliminate all variation. However, once the effects of risk and of statutory interest rate ceilings both have been purged from the data, no significant differential remains between the eastern states and the rest of the country. The mortgage market emerges unscathed from this exercise: the market's adjustment for differences in risk plus the influence of statutory price ceilings together appear to account for the interest differentials present in the Census data.

The Populists would seem to have been right when they asserted that the risks of mortgage lending in the states of the West did not fully account for the existence of regional differentials in mortgage interest rates, but not for the reasons they supposed. Indeed, to the extent that binding interest rate ceilings in the Atlantic states induced mortgage companies to shift funds away from those markets and toward more remunerative loans in the West, farmers in the centers of Populist unrest would have been made better off by the resulting decline in lending rates. Is it possible that western farmers were in fact beneficiaries of the very phenomenon of which they so vociferously complained?

[41] For $1/\phi=1.1$, the presence of a 6 percent ceiling lowers interest rates by .62 of a point, and for $1/\phi=1.2$ by .68 of a point.

REFERENCES

Alston, Lee J., "Farm Foreclosures in the United States During the Interwar Period," *Journal of Economic History*, December 1983, *43*, 885–904.

Argersinger, P. H., *Populism and Politics*, Lexington: University Press of Kentucky, 1974.

Bogue, Allan G., *Money at Interest*, Ithaca: Cornell University Press, 1955.

Bowman, John and Keehn, Richard, "Agricultural Terms of Trade in Four Midwestern States 1870–1900," *Journal of Economic History*, September 1974, *34*, 592–609.

Buck, Solon Justus, *The Granger Movement*, Cambridge: Harvard University Press, 1913.

Carstensen, Vernon, *Farmer Discontent, 1865–1900*, New York: Wiley & Sons, 1974.

Coelho, Philip and Shepherd, James, "Differences in Regional Prices: The United States, 1851–1880," *Journal of Economic History*, September 1974, *34*, 551–91.

Cooley, Thomas F. and DeCanio, Stephen J., "Rational Expectations in American Agriculture, 1867–1914," *Review of Economics and Statistics*, February 1977, *59*, 9–17.

Danhof, Clarence, "Farm-making Costs and the 'Safety Valve': 1850–1860," *Journal of Political Economy*, June 1941, *49*, 317–59.

Davis, Lance, "The Investment Market, 1870–1914: The Evolution of a National Market," *Journal of Economic History*, September 1965, *25*, 355–99.

DeCanio, Stephen, "Economic Losses from Forecasting Error in Agriculture," *Journal of Political Economy*, April 1980, *88*, 234–58.

Eichengreen, Barry, "Currency and Credit in the Gilded Age," in Gary Saxonhouse and Gavin Wright, eds., *Spirit, Technique and Form in the Making of the Modern Econo-*

mies, New York: JAI Press, 1984.

Eliot, Clara, *The Farmer's Campaign For Credit*, New York: Appleton, 1927.

Fogel, Robert and Rutner, Jack, "The Efficiency Effects of Federal Land Policy, 1850–1900: A Report of Some Provisional Findings," in William O. Aydelotte et al., eds., *The Dimensions of Quantitative Research in History*, Princeton: Princeton University Press, 1972.

Frederiksen, D. M., "Mortgage Banking in America," *Journal of Political Economy*, March 1894, *2*, 203–34.

Goodwyn, Lawrence, *Democratic Promise: The Populist Movement in America*, New York: Oxford University Press, 1976.

Hausman, Jerry, "Specification Tests in Econometrics," *Econometrica*, November 1978, *46*, 1251–71.

Hicks, John, *The Populist Revolt*, Lincoln: University of Nebraska Press, 1931.

Higgs, Robert, "Railroad Rates and the Populist Uprising," *Agricultural History*, July 1970, *44*, 291–97.

______, *The Transformation of the American Economy 1865–1914*, New York: Wiley & Sons, 1971.

Hill, F. F., "An Analysis of the Loaning Operations of the Federal Land Bank of Springfield and Its Organization in March 1917 to May 31, 1929," *Bulletin* 293, Cornell University Agricultural Experiment Station, 1932.

Hofstadter, Richard, *The Age of Reform*, New York: Knopf, 1955.

James, John, "Banking Market Structure, Risk, and the Pattern of Local Interest Rates in the United States, 1893–1911," *Review of Economics and Statistics*, November 1976, *58*, 453–62.

Keynes, John Maynard, *The General Theory of Employment, Interest and Money*, London: Macmillan, 1936.

Klepper, Robert W., "The Economic Bases of Agrarian Protest Movements in the United States, 1870–1900," unpublished doctoral dissertation, University of Chicago, 1973.

Lee, Susan Previant and Passell, Peter, *A New Economic View of American History*, New York: W. W. Norton, 1979.

Lee, Virgil P., *Principles of Agricultural Credit*, New York: McGraw-Hill, 1930.

Macauley, Frederick R., *Some Theoretical Problems Suggested by the Movements of Interest Rates, Bond Yields and Stock Prices in the United States Since 1856*, Ann Arbor: NBER University Microfilms, 1938.

McGuire, Robert A., "A Portfolio Analysis of Crop Diversification and Risk in the Cotton South," *Explorations in Economic History*, October 1980, *17*, 342–71.

______, "Economic Causes of Late Nineteenth-Century Agrarian Unrest: New Evidence," *Journal of Economic History*, December 1981, *41*, 835–52.

McMath, Robert C., *Populist Vanguard: A History of the Southern Farmers' Alliance*, Chapel Hill: University of North Carolina Press, 1975.

Mappin, W. F., "Farm Mortgages and the Small Farmer," *Political Science Quarterly*, September 1889, *4*, 433–51.

Mayhew, Anne, "A Reappraisal of the Causes of Farm Protest Movements in the United States, 1870–1900," *Journal of Economic History*, June 1972, *32*, 464–75.

North, Douglass, *Growth and Welfare in the American Past*, New York: Prentice-Hall, 1966.

Robins, Kingman Nott, *The Farm Mortgage Handbook*, New York: Doubleday, 1919.

Severson, R. F., "The Sources of Mortgage Credit for Champaign County, 1865–1880," *Agricultural History*, July 1962, *36*, 150–55.

Shannon, Fred A., *The Farmer's Last Frontier*, New York: Rinehart, 1945.

Skilton, Robert H., "Developments in Mortgage Law and Practice," *Temple University Law Quarterly*, August 1943, *17*, 315–84.

Sparks, E. S., *History and Theory of Agricultural Credit in the United States*, New York: Thomas Y. Crowell Company, 1932.

Stock, James A., "Real Estate Mortgages, Foreclosures, and Midwestern Agrarian Unrest, 1865–1920," *Journal of Economic History*, March 1984, *44*, 89–106.

Summers, Lawrence, "The Nonadjustment of Nominal Interest Rates: A Study of the Fisher Effect," in James Tobin, ed., *Macroeconomics, Prices and Quantities*, Washington: The Brookings Institution, 1983, 201–45.

Sylla, Richard E., *The American Capital Market*

1846–1914, New York: Arno, 1975.

Trojanowski, Joseph, "The Stability of Freight Rate Agreements at Minor Railroad Junctions in Iowa and Kansas, 1880–1910," unpublished doctoral dissertation, Yale University, 1980.

Unger, Irwin, *The Greenback Era*, Princeton: Princeton University Press, 1964.

Williams, Jeffrey C., "Economics and Politics: Voting Behavior in Kansas during the Populist Decade," *Explorations in Economic History*, July 1981, *18*, 233–56.

Williamson, Jeffrey G., *Late Nineteenth Century American Development*, Cambridge: Cambridge University Press, 1974.

Wright, Gavin, *The Political Economy of the Cotton South*, New York: W. W. Norton, 1978.

Wright, Ivan, *Farm Mortgage Financing*, New York: McGraw-Hill, 1923.

Zartman, Lester, *Investments of Life Insurance Companies*, New York: H. Holt, 1906.

Illinois State Bureau of Labor Statistics, *Fifth Biennial Report*, Springfield: Springfield Printing Co., 1888.

U.S. Bureau of the Census, (1895a) "Report of Real Estate Mortgages in the United States," in *Eleventh Census of the United States*, Vol. XII, Washington: USGPO, 1895.

______, (1895b) "Report of the Statistics of Agriculture in the United States," Vol. III, Washington: USGPO, 1895.

U.S. Comptroller of the Currency, *Annual Report*, Washington: USGPO, various issues.

U.S. Department of Commerce, *Historical Statistics of the United States*, Washington: USGPO, 1975.

Wisconsin State Bureau of Labor, Census and Industrial Statistics, *Seventh Biennial Report*, Madison: Democratic Printing Company, 1896.

Applied General Equilibrium Analysis of Small Open Economies with Scale Economies and Imperfect Competition

By RICHARD HARRIS*

This paper outlines the theoretical structure and empirical implementation of a real trade general equilibrium model of a small open economy incorporating some features associated with the "industrial organization" (*IO*) approach to trade. While the theoretical literature[1] in this area is quite recent and growing rapidly, the general concern with imperfect competition, economies of scale, entry barriers, product differentiation, and other aspects of industry structure is certainly not new to the discussion of trade economists on the costs and benefits to trade liberalization. Bela Balassa (1966), W. M. Corden (1972; 1974), H. C. Eastman and S. Stykolt (1966), and Ron Wonnacott and Paul Wonnacott (1967) have all emphasized the role of scale economies and its impact on trade and industrial structure in small open economies. A related theme is the observation of Balassa (1966) and H. G. Grubel and P. J. Lloyd (1975) that much trade occurs on an intra-industry basis and this provides scope for intra-industry adjustment, as well as the interindustry adjustment associated with the Hecksher-Ohlin view of comparative advantage. For small open economies such as Australia, Canada, and Sweden, there is a long history in the policy debate over free trade as to the adverse consequences of protection on efficiency and factor productivity in manufacturing industries for similar reasons.[2]

The common argument put forth in the *IO* view is that protection, by restricting market size and limiting foreign competition, promotes too many firms within an industry, operating at too small a scale, and with too many product lines being produced within the plant. Related is the observation that protection may facilitate oligopolistic coordination of the protected firms.[3] It is well known that conventional calculations of the costs of protection give numbers which are quite small; often in the order of 0.5 to 2 percent of *GNP*. This result holds for almost all known studies based on the competitive neoclassical model, either partial or general equilibrium. Recent examples include Robin Boadway and John Treddenick (1978), Fred Brown and John Whalley (1980), W. R. Cline et al. (1978), A. V. Deardorff and R. M. Stern (1981), P. Dixon et al. (1981), Steven Magee (1972), and J. Williams (1976). The remarkable robustness of these small welfare gain estimates has been of little comfort to economists promoting free trade. On the other hand, those of the *IO* persuasion have suggested these estimates are too small, particularly for small open economies, because of the competitive constant returns assumptions underlying the analysis. Balassa (1966; 1975) and Wonnacott-Wonnacott argue the costs of protection are much higher than suggested by conventional "triangle" analysis. A recent paper by W. D. Ingram and

*Queen's University, Kingston K7L 3N6, Canada.
The research support of the Ontario Economic Council is gratefully acknowledged. The views contained in this paper are mine alone and do not in any way reflect the policies or position of the Ontario Economic Council. Earlier versions of this paper were presented at the University of Western Ontario and 1982 European Econometric Society Meetings, Dublin. Comments of the discussants and participants in those sessions and the suggestions of an anonymous referee are gratefully acknowledged. I also thank David Cox for numerous insightful observations throughout the course of the project from which this paper derived. Remaining errors are my responsibility.

[1] The relevant literature will be cited later in this and subsequent sections.

[2] Much of the policy literature relating scale economies and imperfect competition to trade liberalization is cited in the survey by Corden (1976). For Australia the policy discussion is summarized by Lloyd (1978) and for Canada in the Economic Council of Canada (1975).

[3] See Eastman and Stykolt.

S. R. Pearson (1980) finds large gains to the formation of a customs union between Ghana and the Ivory Coast through incorporating scale economies.

While the cost of protection question has received the most attention, the industrial organization perspective is significant for any positive or normative questions in international trade if the determinants of trade are thought to be dependent upon *IO* features of the economy. By its very nature the *IO* perspective also bears significantly on questions of industrial policy and antimonopolies policy in small open economies.

Trade theory and commercial policy is a branch of economics which is inherently concerned with general equilibrium questions. The extensive development of the Hecksher-Ohlin model of trade has proved to be a powerful tool for economists. Industrial organization on the other hand, has long been an area where the method of partial-equilibrium analysis is predominant. Bringing the two together is not easy. The recent theoretical work of James Brander (1981), Elhanan Helpman (1981), Paul Krugman (1980), and Kelvin Lancaster (1980) has been a major advance in this respect. The integration is inevitably difficult, and extremely complex. A critic might argue that the features of trade which are explained by incorporating industrial organization details are not quantitatively significant. What is significant are the general equilibrium intersectoral relations between commodity flows and goods and factor prices. The theoretical complexity and untidiness of incorporating industrial organization within a general equilibrium framework is simply not worth the effort.

It is my intention to demonstrate that this view is incorrect on two counts. First, it is possible to construct an *empirical* general equilibrium model of a small open economy which captures many of the industrial organization features of the industries in the economy thought to be important. The model while complex is not unduly so, and is implementable given existing data sets for modern developed economies. The empirical general equilibrium methodology used here is in the spirit of that used by John Shoven and Whalley (1983).[4] Using a variety of data sources and econometric estimates, the model is required to produce an observed historical data set as an equilibrium. In this paper I will not detail this procedure or document the data sources. For the interested reader this is contained in my 1984 study.

Second, it will be shown that the quantitative and qualitative significance of incorporating industrial organization features in a general equilibrium trade model are considerable. This has long been suggested to be the case in the costs of protection debate, as noted above. The quantitative results of a general equilibrium analysis that includes *IO* features differs significantly in many cases from one that does not. To give one example, the estimated static long-run gains to Canada of free trade are in the range of 8–12 percent; considerably larger than suggested by conventional methods.

It is important to emphasize that the scale economies referred to in this paper are at the level of the individual plant in the manufacturing sector and hence *internal* to the firm. Because they are internal to the firm, the industries in question are *necessarily* imperfectly competitive. There is a tradition in the theoretical trade literature of referring to scale economies as external to the firm but internal to the industry; this allows one, of course, to maintain the assumption of perfect competition. The more recent theoretical work in trade by Helpman and Krugman is explicitly concerned with the interaction between scale economies internal to the firm and industrial organization. Internal scale economies at the plant level is the feature of industry given the most emphasis in the Canadian policy debate.[5] While external scale economies may be significant, little is known about their empirical magnitude. Fortunately, however, the empirical industrial organization literature tells us a lot about scale economies internal to the firm. It is

[4] Shoven and Whalley is a recent survey of their own and others' work in the applied general equilibrium area.

[5] A summary of the Canadian trade policy debate and references to the appropriate literature can be found in the Economic Council of Canada.

these empirical estimates which form a crucial input to the model.

Given the existence of internal scale economies and hence the necessity for an imperfect competition view of industry structure, one is left with a wide variety of potential market structures which may be appropriate in any particular empirical context. In this paper a fairly conventional free-entry Chamberlinian industry structure is used. Free entry, together with the implication of price equal to long-run average cost, is probably the most important feature of the model. It is well known, though, that free entry alone does not determine individual firm output and hence average cost. The manner in which prices are set within the industry, together with the degree of scale economies and free entry jointly determine industry price, output, and cost. At one extreme, prices can be set in a noncooperative Cournot-Nash fashion as in the monopolistic competition model. At another extreme, prices can be set collusively at some appropriate focal point. In the trade literature, one hypothesis that has received a great deal of attention is that put forward by Eastman and Stykolt. Collusive domestic oligopolies use the world price plus tariff as a focal point at which to set industry price. Little direct empirical evidence is available which discriminates between these two hypotheses. In the model, prices are set by taking an average of the two prices. An industry that is *more* competitive might be characterized as one in which prices are set closer to the monopolistically competitive price than the collusive price. In commenting upon the empirical results of the model, some attempt will be made to report to what extent the results are sensitive to the degree of imperfect competition.

The conclusion of the paper is that industrial organization features *matter* in the discussion of trade within a small open economy. Furthermore, they matter even more when embedded within a general equilibrium framework, in which intersectoral resource shifts and factor price changes are accounted for. Indeed, the results suggest a stronger conclusion. To ignore these features, if it is known a priori they are significant at the industry level, is to seriously misspecify the general equilibrium analysis. The misspecification relates both to predictions at the industry *and* the aggregate level. Admittedly the quantitative results are suggestive given the model to date has been implemented on only one data set. Nevertheless, they are striking.

I. An Overview of the Model

The model is a real trade model of a small open economy with one fixed factor, labor, and one internationally mobile factor, capital services; both are mobile between industries and firms. Industries divide a priori into those which are competitive and constant cost, and those which are noncompetitive increasing returns industries. There are two versions of the model; one in which there is no product differentiation (non-*PD*) and one in which there is (the *PD* model). While both models allow scale economies, the *PD* model also allows for multiproduct firms choosing the set of products to produce. In this case there are decreasing costs at the level of the plant for a given set of product lines, and there are economies of scale to longer production runs, or equivalently diseconomies to having a larger number of product lines within a given plant. In each noncompetitive industry, with product differentiation, products produced by different firms are imperfect but close substitutes. Likewise, foreign and domestic goods in any industry are imperfect substitutes. For this and other reasons (see Brander) intra-industry trade is a characteristic of the model.

The economy is presumed to be a price taker in its import markets and a price maker in export markets; an "almost" small open economy (*SOE*). This is a fairly standard assumption and is consistent with the evidence for Canada. Domestically produced goods are used in both final demand (domestic and foreign) and as intermediate inputs. All factor markets are competitive with rental capital in infinitely elastic supply at the world rental rate. In competitive industries, constant returns is assumed so price must equal unit cost.

Much of the novelty of the model is with respect to the treatment of the firm in noncompetitive industries. In the *PD* model, all products within an industry are viewed as

symmetric imperfect substitutes. In both models, all firms have access to the same technology, and each industry consists of some endogenous number of representative firms. The firm produces a set of representative industry products, and chooses a representative price for each product. Quantity sold is demand determined once the firms set price. The firm sets its price by marking up on unit variable cost. Two alternative pricing hypotheses are used. One uses a Takashi Negishi (1961) perceived demand curve approach. The firm has a perceived constant elasticity demand curve which determines the optimal markup (monopolistic pricing hypothesis: *MPH*). The other is a type of collusive pricing hypothesis proposed by Eastman and Stykolt. The domestic price within an industry is set equal to foreign price of the industries import competing good plus tariff as a collusive focal point. In the simulations, actual price is set by taking a weighted average of the monopolistic price and Eastman-Stykolt price.

In the *PD* model, it is assumed foreign importers match in percentage terms any changes in domestic product differentiation. This assumption of competitive foreign product differentiation (*CFPD*), or something analogous to it is required in a *SOE* model with an otherwise exogenous foreign sector. The reason is that increasing domestic relative to foreign product differentiation squeezes out foreign imports at constant industry terms of trade.

Firms enter and exit in response to pure profits and losses (free-entry/exit assumption). The only entry barriers are fixed costs. Equilibrium at the industry level requires that all firms make (approximately) zero profits. Under *MPH*, a "true elasticity" and perceived elasticity are required to be equal. With *PD*, firms also choose the profit-maximizing number of product lines. This type of equilibrium at the partial-equilibrium level is analogous to the Cournot-Chamberlin equilibrium.

A general equilibrium results when all industries are in equilibrium, all product and factor markets clear, and the balance of payments is in equilibrium. Balance of payments equilibrium is current account balance, or trade surplus equal to rental payments on foreign-owned capital. All tax and tariff revenue is returned to consumers in a nondistorting manner.

A. *The Model*

In the following subsections some details of the model are presented. For the most part, the *PD* and non-*PD* models are similar, and the differences will be pointed out where relevant. Since the model is a multisector-multicommodity, it is not intended to provide theoretical propositions. Nevertheless, given the relative novelty of this type of approach some of the relevant forces at work in the model will be discussed. As in the Helpman model, there are both comparative advantage or interindustry effects, and intraindustry or industrial organization effects present. In a general equilibrium framework neither is independent of the other.

Table 1 lists some of the relevant notation used in this section. The model will be described without taxes, tariffs, or subsidies. In the empirical implementation of the model, most of the relevant tax and tariff distortions are present.[6] The reader interested in the complete set of equations in the model is referred to my earlier study.

B. *Technology*

In this subsection the technology of firms is described in detail. Costs are divided into fixed and variable costs. In the competitive industries there are no fixed costs. Total variable costs are derived from a constant returns to scale production function. Denote the unit variable cost function as $v^i(P)$. The particular form used for v^i is the Cobb-Douglas.

$$\text{(1)}\quad \log v^i(P) = \alpha_{0i} + \sum_{j \in M} \alpha_{ij} \log w_j^i + \sum_{k \in B} \alpha_{ik} \log q_k + \alpha_{i_w} \log w + \alpha_{i_r} \log r.$$

[6] My earlier study contains the details of the relevant tax and tariff distortions in the empirical implementation of the model.

TABLE 1—SOME NOTATION

N: index set	for noncompetitive industries
C: index set	for competitive industries
B: index set	for noncompeting imports
Other index sets	$M = N \cup C$; $G = M \cup B$
$(p_i)_{i \in M}$	domestic commodity prices
$(p_i^*)_{i \in M}$	foreign commodity prices
$(q_i)_{i \in B}$	foreign prices on noncompeting imports
w:	domestic wage
r:	world rental rate on capital
$P =$	(p, p^*, q, r, w): price system
D_i:	domestic product set for $i \in N$
n_i:	number of products in D_i
D_i^*:	foreign competing product set, $i \in N$
n_i^*:	number of products in D_i^*

The parameters $\{(\alpha_{ij})_{j \in M, B}, \alpha_{iw}, \alpha_{ir}\}$ and $\{\beta_{ij}\}_{j \in M}$ are conventional share parameters from a constant returns Cobb-Douglas production function; α_{iw} and α_{ir} are the share parameters on the primary inputs, labor and capital, respectively; w_j^i is a price index of a composite input produced by industry j, used by industry i, a composite of both foreign and domestic inputs from j. The index w_j^i is assumed to have the form

$$(2) \quad \log w_j^i = \beta_{ij} \log p_j + (1-\beta_{ij}) \log p_j^*.$$

Equation (1) assumes that foreign and domestic intermediate inputs have a unitary elasticity of factor substitution; it also implies that the degree of product differentiation in each industry has no effect on intermediate demand for the composite input. The input-output matrices for the economy are derived from the unit cost functions assuming price-taking behavior in input markets. The domestic Leontief matrix $[a_{ij}(P)] \equiv A(P)$ is defined by

$$(3) \qquad a_{ij}(P) = \alpha_{ji}\beta_{ji} v^j(P)/p_i,$$

where a_{ij} is the demand for domestic composite good i, per unit output of composite domestic good j. The Leontief matrix for foreign imports $A^*(P)$ is defined as

$$(4) \quad a_{ij}^*(P) = \alpha_{ji}(1-\beta_{ji}) v^j(P)/p_i^*.$$

Each representative firm in each noncompetitive industry, $i \in N$, has both fixed and variable costs. Fixed costs are given by the function

$$(5) \quad F_i(r, w, k_i) = rf_K^i + wf_L^i + (rh_K^i + wh_L^i)(k_i - 1)^{\gamma_i}, \quad k_i \geq 1.$$

The fixed costs of any firm include *plant* fixed costs and product fixed costs. To set up a plant requires using f_K^i units of capital; this is the amount of capital required to produce the minimum number of product lines. k_i, the number of product lines has a minimum at $k_i = 1$. Plant fixed costs also include wf_L^i in labor costs; this can be thought of as the fixed labor costs necessary to keep the plant open. The second part of the cost function represents the amount of *product specific* capital and labor required as the number of product lines increases beyond the minimum. Product specific costs include the additional capital necessary in a horizontally diversified plant plus the additional labor costs associated with product changeovers.[7]

In the non-*PD* model, all domestic firms are assumed to produce only one homogeneous good. Thus all firms produce with k_i set equal to one and the distinction between product and plant fixed costs is unnecessary.

Define z_i as the length of the production run on a representative product in the representative firm in industry i. Total composite "output" in the firm is defined as $y_i = k_i z_i$.[8]

[7]In a true dynamic model there is a production period, endogenously chosen, and a number of product changeovers, together with lengths of production runs, all choice variables on each product. Armen Alchian (1959) emphasized the tradeoffs between these different choice variables within the firm. In the model of this paper the production period is the same as basic economic period over which all flows are measured—it is not a choice variable. Second, the cost function does not distinguish between plants which undertake their production of alternative products sequentially, one after another throughout the period, and plants which simultaneously produce all products. Clearly in practice there are important differences in alternative methods of organization.

[8]In a multiproduct firm there is, not surprisingly, an index number problem associated with defining firm "output." The multiplicative index chosen here is *ad hoc*, but has the virtue that it is analytically simple, easy

Total costs in the firm are

$$(6) \qquad TC_i = v^i(P)y_i + F_i(r,w,k_i).$$

Let $AC^i(y_i|k_i)$ be average cost per unit output *given* a fixed number of product lines. Let $DC^i(k_i|\bar{y}_i) = v^i + F_i(k_i)/y_i$ be average cost per unit of output as the number of product lines vary, holding total output constant. Figure 1 illustrates the relevant assumptions regarding these costs. While k_i should be treated as an integer variable, it does little harm for empirical purposes to treat it as continuous. For practical purposes, maximum economies of scale (MES) output is defined as the level of y_i where average cost is within 1 percent of average variable cost. As in the monopolistic competition model, firms operate at MES or below. In free-entry equilibrium, low-fixed-cost industries will be relatively competitive with small MES and large numbers of firms. Formally the technology of the multiproduct firm as specified in (6) embodies "economies of scope."[9]

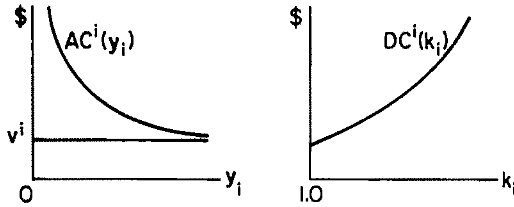


FIGURE 1

C. *Final Demand*

Final demand in each commodity category consists of export demand by rest of world (ROW) and final domestic demand. In equilibrium, given PD, domestic industry i produces n_i products; each *product* with an export demand in amount e_i. To keep the model empirically tractable, it is necessary that the export demand equation be specified at the industry level, rather than individual product level. Let E_i denote industry i exports defined as $E_i \equiv n_i e_i$. The demand per product, e_i, is assumed to depend inversely on the total number of products sold, n_i. This might be thought of as a unit elastic product "crowding out" effect. It implies that the total export demand, E_i, can be taken as independent of the level of industry product differentiation.

There is a particular difficulty in specifying export demand equations which relates to the interpretation of the ROW. In reality, the ROW is not a single country but many countries, and one country's exports compete with many other countries' exports in the "world" export market. Without building a world model it is very difficult to adequately capture this in a reduced-form export demand equation. The following might be viewed as a "half-way" approach.

The ROW is thought of as consisting of many countries both exporting and importing. In any particular commodity category there is a world demand for a composite export good, which is an aggregate of all relevant countries' exports. Dropping subscripts for convenience, the composite export good is defined as

$$(7) \quad Q = \left[\beta E^{-\lambda} + (1-\beta)E^{*-\lambda}\right]^{-1/\lambda},$$

where E is domestic exports, and E^* is all other countries' exports; $\sigma^* = 1/(1+\lambda)$ is the elasticity of substitution between domestic and other exports in world demand. Letting p and p^* be the price paid by world consumers of domestic and other exports, the price index for the export composite is

$$(8) \quad \hat{P} = \Big[\beta^{\sigma^*} p^{(1-\sigma^*)} + (1-\beta)^{\sigma^*} p^{*(1-\sigma^*)}\Big]^{1/(1-\sigma^*)}$$

to interpret, and conforms to commonly used empirical measures of plant output in scale economy studies.

[9]On economies of scope, its formal definition and implications, see John Panzar and Robert Willig (1981). Economies of scope here means that a firm can lower its average cost on the average product by producing a diversified output vector, taking into account the possibility of replicating plants. Thus, although there are economies to specialization within the plant, over a range of product diversity, average costs are lower by producing multiple products within a single plant; this avoids incurring the additional fixed costs associated with replicating plants.

For a given Q, cost minimization yields the demand for domestic exports

$$(9) \qquad E = \beta^{\sigma^*}(\hat{P}/p)^{\sigma^*} Q.$$

A conventional constant elasticity demand function for the ROW composite export good is assumed:[10]

$$(10) \qquad Q = \alpha \hat{P}^{-\varepsilon}.$$

Substituting (8) and (10) into (9) gives the demand for domestic exports

$$(11) \quad E = \alpha' \left[\beta^{\sigma^*} p^{(1-\sigma^*)} + (1-\beta) p^{*(1-\sigma^*)} \right]^{\frac{\sigma^* - \varepsilon}{(1-\sigma^*)}} / p^{\sigma^*},$$

where $\alpha' = \alpha \beta^{\sigma^*}$. The importance of this specification is in its implication for price and foreign tariff elasticities. Let t^* denote the world ad valorem tariff rate against domestic and other exports. Then

$$\partial \log E / \partial \log(1 + t^*) = -\varepsilon,$$

and

$$\partial \log E / \partial \log p = -\left[\theta \varepsilon + (1-\theta) \sigma^* \right] \equiv \mu,$$

where θ is the domestic value share of world exports, $p(1+t^*)X/\hat{P}Q$. Thus, price elasticities of domestic exports depend not only on the world export elasticity, but also on the extent to which domestic and other goods are viewed as substitutes as reflected in σ^*. For example, it is quite conceivable that the aggregate export elasticity, ε, is a reasonable value, -3, but a particular country's export price elasticity in absolute value is quite low because its share of the world export market is low and σ^* is reasonably high, say 2. This would typically be the case for example of industries which are thought to be import competing. The industry export demand equation (11) is the one used in the model for both the PD and non-PD cases.

Domestic final demand is generated by treating the economy as a single consumer. Adding multiple consumers would add little to the IO aspects of the model but would be important if interested in the incidence of the costs of protection across different income classes. The aggregate consumer maximizes a utility function over all commodity categories of the log-linear form

$$(12) \qquad \log U = a_0 + \sum_{i \in \boldsymbol{G}} a_i \log C_i.$$

For noncompeting imports, $i \in \boldsymbol{B}$, C_i represents the amount of the import good. For $i \in \boldsymbol{C}$, competitive industries, the "Armington" assumption is used; foreign and domestic goods are imperfect substitutes as reflected in the CES aggregator,

$$(13) \qquad C_i = \left[\delta_i x_i^{\rho_i} + (1-\delta_i) x_i^{*\rho_i} \right]^{1/\rho_i}$$

where x_i and x_i^* are domestic and foreign goods, respectively, in final demand.

In PD industries, a Spence-type generalization (see A. Michael Spence, 1976) of (13) is used which allows the product set to vary.

$$(14) \qquad C_i = \left[\sum_{v \in D_i} x_{iv}^{\rho_i} + \sum_{v \in D_i^*} x_{iv}^{*\rho_i} \right]^{1/\rho_i}.$$

The common elasticity of substitution between all goods v in category i product set, $D_i \cup D_i^*$, is $\sigma_i = 1/(1-\rho_i)$ for $i \in \boldsymbol{N}$, $\sigma_i > 1$.[11] In equilibrium all $x_{iv} = x_i$ and all $x_{iv}^* = x_i^*$.

Given disposable income Y, and prices P, the demand for a representative product in industry i by the consumer is

$$(15) \qquad x_i = a_i Y p_i^{-\sigma_i} / \left[n_i p_i^{-\sigma_i + 1} + n_i^* p_i^{*-\sigma_i + 1} \right]$$

[10] H. D. Evans (1972) is an early example of the use of export demand equations such as (10) in applied general equilibrium work.

[11] We require $\sigma_i > 1$ because this is approximately the absolute value of the own-price elasticity of demand for an individual product in category i. With a Cobb-Douglas specification the own-price elasticity for a commodity aggregate is -1. Aggregating goods into a particular category only makes sense if they can be viewed as closer substitutes within a category than between categories.

Multiplying (15) by the number of domestically produced goods, n_i, gives total domestic final demand for the domestic composite i, X_i,

$$(16)\quad X_i = n_i x_i = a_i Y n_i p_i^{-\sigma_i} / \left[n_i p_i^{-\sigma_i+1} + n_i^* p_i^{*\,-\sigma_i+1} \right].$$

In (16), the relative number of domestic to foreign imported goods, (n_i/n_i^*) acts like a share parameter in the conventional *CES* function. Consequently, at constant terms of trade the domestic industries share in domestic consumption is increasing in the ratio n_i/n_i^*. The subutility function (14) incorporates the Chamberlin effect of a taste for diversity on the part of consumers. The utility specification (14) is subject to the criticism that the elasticity between products is independent of the number of products, and, furthermore, consumers will always consume any number of goods, however large, provided they are available at finite prices. The "elasticities approach" to product differentiation may therefore overstate the benefits to consumers of product differentiation.

The degree of competing foreign product differentiation, n_i^*, should properly be treated as an endogenous variable. The assumption employed is that importers match in proportional terms any changes in domestic product differentiation. This hypothesis, referred to as competitive foreign product differentiation, keeps (n_i/n_i^*) constant. This hypothesis on the behavior of importers seems reasonable as it keeps foreign and domestic market shares constant, given constant industry terms of trade. Other assumptions are plausible. A more elaborate model might attempt to explicitly model foreign importers.

D. *Short-Run Equilibrium*

The short run is a period in which industry *structure* is fixed. The following variables are held constant in the noncompetitive industries: markups on unit cost by firms, $(m_i) = m$; number of firms in each industry, $(Fm_i) = Fm$; number of product lines of each firm, $(k_i) = k$; and number of domestic and foreign product lines n and n^*. Let $S = (m, Fm, k, n, n^*)$ be the vector of industry structure variables. All other economic variables adjust within the short run. This includes commodity and factor prices, outputs, employment of variable factors, etc. This short run is similar to but not quite the same as the Marshallian short run of textbook economics.

In the short run, pure profits and losses will occur in each of the noncompetitive industries. Let π_i denote industry i's pure profit or loss after paying all factors opportunity cost. Aggregate consumer income is given by

$$Y = wL + rK_D + \Psi \sum_{i \in N} \pi_i,$$

where L is the aggregate labor endowment, K_D the domestic capital endowment, and Ψ the share of domestic ownership in industry; these are taken as exogenous (even in the long run) variables.

Equilibrium commodity prices are determined by the equations

$$(17)\quad p_i = m_i v^i(P), \qquad i \in N$$
$$p_i = v^i(P), \qquad i \in C.$$

Solving equation (17) will determine domestic commodity prices as a function of factor prices, (w, r), and world prices, $(p_i^*)(q_i)$. Let $X(P, Y, S)$ and $E(P)$ denote final demand vectors. Commodity and labor market clearing implies the vector of total outputs, Z, must satisfy,

$$(18)\quad Z = \left[I - A(P)^T \right]^{-1} \left[X(P, Y, S) + E(P) \right]$$

$$(19)\quad L = \sum_{i \in M} a_{iw}(P) Z_i.$$

A short-run equilibrium for a given S is a wage $w(S)$, domestic commodity price vector $p(S)$, income $Y(S)$, and vector of outputs $Z(S)$ satisfying (17)–(19). This definition has to be modified with the addition of

taxes and tariffs in the usual way. Walras' law implies a balance of payments equilibrium of the form

$$(20)\quad \sum_{i \in M} p_i E_i - \sum_{i \in G} p_i^* M_i = r(K - K_D) + (1 - \Psi) \sum_i \pi_i;$$

surplus on trade account equals rental payments on net capital service imports, $r(K - K_D)$, plus quasi rents to foreign ownership, $(1 - \Psi)\pi$; that is, current account balance. K is total domestic demand for capital services.

With PD, the following identities hold:

$$n_i = Fm_i k_i; \qquad y_i = Z_i / Fm_i.$$

In non-PD equilibrium n_i, n_i^*, and k_i are treated as constants in both the short and long run.

E. *Firm Behavior*

I now turn to how the firm behaves with respect to its pricing and product decisions. Under the monopolistic pricing hypothesis (MPH), each firm in each industry has a perceived demand curve for its representative product, holding the number of products fixed and the price of its competitors' products constant. Let z_{if} be the perceived demand in industry i, by firm f for its representative product. The *perceived* demand curve for this product is assumed to have the constant elasticity form

$$(21)\qquad z_{if} = \Psi_{if} p_{if}^{-\varepsilon_{if}}.$$

The optimal pricing rule, given (22) is the Lerner formula

$$(22)\qquad (p_i - v_i)/p_i = 1/\varepsilon_i,$$

where the firm subscript has been dropped for convenience. It remains to determine where the elasticity ε_i comes from.

Under the Eastman-Stykolt hypothesis (ESH), the firm sets its price equal to the import competing goods price, inclusive of the domestic tariff, $t_i p_i^*$:

$$(23)\qquad p_i = p_i^*(1 + t_i).$$

As discussed earlier this is a type of collusive pricing policy which might be expected in protected oligopolistic industries. In many of the policy simulations discussed, a mixed pricing hypothesis is adopted in which the price set is an average of the MPH and ESH prices.

In the case of product differentiation it remains to determine the number of product lines in each firm. Let $z(p, k)$ denote the *perceived demand curve* facing a firm for a representative product given its price p, and a number of product lines k. The firm must solve the problem

$$(24)\qquad \max_{p,k} (p - v) z(p, k) k - F(k)$$

where $F(k)$ is total fixed costs as a function of k. To make the problem tractable the individual firm in each industry is assumed to take total industry demand Z as given and unchanged by changing its number of products, given all prices are held constant. The firm observes its current demand per product $\hat{z} = \hat{Z}/\hat{n}$, and current level of industry product differentiation $\hat{n}$. Given its current number of product lines $\hat{k}$, it contemplates adding Δ product lines. Its approximate problem then is to choose Δ to maximize

$$(25)\quad (p - v)(\hat{k}\hat{Z}/(\hat{n} + \Delta) + \Delta z) - F(\hat{k} + \Delta)$$

In equilibrium, $k = \hat{k}$, or $\Delta = 0$, and the first-order condition is

$$(26)\quad (p - v) y (1 - 1/Fm) = F'(k) k.$$

Clearly more sophisticated hypotheses as to the firms' potential product demand are possible, but this one seems informationally realistic.

An implication of (26) is that for industries with a very large number of firms $(p - v)z = F'(k)$. The zero profit condition (see subsection F below) is $(p - v)z =$

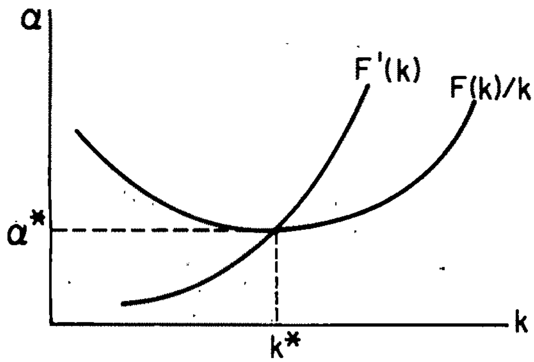


FIGURE 2

$F(k)/k$. Letting $(p-v)z \equiv \alpha$, (26) and the zero profit condition jointly determine (α^*, k^*) as in Figure 2. Analogous to the Marshallian analysis of firm output, k^* is determined independently of α, and given by the intersection of average and marginal *fixed* cost curves. Thus the only thing that shifts k^* will be changes in the fixed cost technology or factor prices. The composition of α^* between output and markup is determined essentially as in the Chamberlin model and is closely related to the perceived elasticity of demand. The more elastic the perceived demand curve, the lower the markup and hence, for a given α^*, the greater the output from the plant on any *given* product. Changes in markups are an important source by which the benefits from scale economies are achieved.[12]

[12] In this model the maximum benefits from product rationalization within the plant occur at that k^* at which the total fixed costs are minimized in providing some aggregate number of products. For example, suppose an omniscient planner wants to deliver N products, all in quantities z. To minimize the aggregate cost of doing so he would solve $\min_k Nvz + (N/k)F(k)$. It is clear he would want to minimize average fixed costs per product, or choose $k = k^*$. The upshot of this discussion then is that large numbers competition and zero profits, given the decision rule (26), result in an approximately efficient degree of horizontal product diversity within the firm. With small firm numbers or nonzero profits, this will not be true.

F. *Entry, Perceived Demand, and Long-Run Equilibrium*

To close the model it is assumed firms enter and exit in response to the presence of pure profits and losses as in the classic Marshallian adjustment process. A long-run equilibrium is a short-run equilibrium with two additional conditions.

(i) All industries are in (approximately) a zero pure profit condition.

(ii) Under *MPH*, the perceived elasticity is the "true" elasticity.

What is the true demand curve of the firm? Condition (ii) requires that the firm be *locally* correct in its perceptions as to its true demand curve. There are clearly different routes one can go as to the true demand curve.

Begin with the identity for total industry output:

$$(27) \qquad Z_i = X_i + E_i + \sum_{j \in M} I_{ij},$$

where $I_{ij} = a_{ij} Z_j$, intermediate use of commodity i by industry j. The elasticity of Z_i with respect to a variable x is[13]

$$(28) \quad \eta_i(Z_i) = (X_i/Z_i)\eta_x(X_i) + (E_i/Z_i)\eta_x(E_i) + \sum_{j \in M} (I_{ij}/Z_i)\eta_x(a_{ij}Z_j).$$

The firm, in calculating its price elasticity of demand, does a general equilibrium comparative statics exercise, changing its own-price and taking as constant (i) all other prices, (ii) degree of product differentiation, (iii) aggregate income, and (iv) the output of other industries. The last is the most crucial. It implies the firm does not account for induced changes in intermediate demand as a result of changes in the marginal cost of production to other industries. Consequently, the elasticity of intermediate demand arises solely from the factor substitu-

[13] $\eta_x(Z)$ denotes the elasticity of Z; with respect to a change in x.

tion effect. An alternative to this would be to differentiate (27) totally, but this is an extremely complex calculation. The important point to note is that shifts in the composition of total demand will change the elasticity of the demand curve and hence the optimal markup.

In going from the industry demand curve to the firm's demand elasticity, the formulae are essentially the same given the assumption of symmetric substitutes, and assuming all firms share the same proportions of each demand category.[14] In the non-*PD* model, the firm's perceived elasticity obviously depends on the nature of price competition among domestic firms. One easy way to characterize this is by the conjectural variation held by each firm on its competitors' reactions, a commonly used device in industrial organization. The problem is that the conjectural variation is usually chosen arbitrarily. Given that one is willing to pick a value for the conjectural variation, it is straightforward to calculate equilibrium. Rather than doing this, and in order to preserve some comparability for the quantitative results reported in the following section, the perceived elasticities in the non-*PD* model are assumed to be the same as the elasticities in the *PD* model. The *ESH* pricing is quite naturally the same in both models. Tacit collusion among domestic firms is assumed and the foreign good(s) are presumed to be close substitutes for the domestic good(s).

The model is general equilibrium with imperfect competition and nonconvexities in production. For the usual reasons, neither existence nor uniqueness of equilibrium is guaranteed. In practice, the algorithm used has always found an equilibrium given the parameter values used. Nonuniqueness has not been encountered by using alternative starting values in the solution algorithm.

III. Welfare and Resource Allocation

In the models specified in Section II there are a number of features that differentiate them from the usual constant returns neoclassical model. This in turn implies that any change in exogenous variables, policy-induced or otherwise, the pattern of resource reallocation and the welfare implications of the reallocation can in principle be quite different than in the usual model. First, in terms of the pattern of resource reallocation, the presence of scale economies implies that there is scope for intra-industry changes in the pattern of production. A given volume of domestic production can be spread amongst a larger or smaller number of firms and therefore change the absolute efficiency of resources utilized in that industry. The endogenous nature of this process works through changes in markups on variable costs and the subsequent entry or exit of firms in response to profits or losses. Lower markups, together with a zero profit free-entry equilibrium, imply for a given cost function that output per plant must be greater and hence average costs lower. Consequently, a primary transmission mechanism in the model occurs through the change in markups in the noncompetitive industries. In the model, markups can fall either because of a fall in the collusive price due to a fall in the border price of competing imports under the *ESH* pricing rule, or because of a shift in the composition of industry demand toward those demands which are relatively more elastic. Interindustry shifts in resources will depend not only on relative price changes and relative factor intensities, but also changes in the

[14] The demand elasticities under *PD* are as follows:

(final) $$\eta_{p_i}(X_i) = -\sigma_i + s_i(\sigma_i - 1)/Fm_i$$

(exports) $$\eta_{p_i}(E_i) = \mu_i$$

(intermediate) $$\eta_{p_i}(I_{ij}) = \tau_{ij} + \tau_{ij}\alpha_{ij}\beta_{ij}/Fm_i - 1,$$

with $s_i = p_i X_i / a_i Y$. The elasticity of intermediate demand introduces the parameters τ_{ij}. In each industry j the *perceived* input from domestic i is taken as a *CES* aggregate with elasticity of substitution τ_{ij}. This yields the output constant factor-demand elasticity above. In the true model, however, the aggregator is Cobb-Douglas with share parameters $\beta_{ij}\alpha_{ij}/n$ on each domestic product in category j. The difficulty with maintaining the Cobb-Douglas assumption is that intermediate demand elasticities are very close to -1; this yields unreasonably high markups. In the simulation model the τ_{ij} are chosen in the range 3.0 and 5.0.

relative efficiency of factor use between industries.

For similar reasons there are additional avenues, other than the usual static production and consumption gains, through which aggregate real income can change in response to an exogenous change. These include: (a) realization of economies of scale at the plant level; (b) longer production runs within the plant; (c) aggregate efficiency gains due to intersectoral resource shifts as a consequence of (a) and (b); and in the case of product differentiation, (d) increases in real utility due to increased domestic and foreign product differentiation.

Theoretically, it is difficult in a general equilibrium framework to make any qualitative predictions as to the effects of any policy changes. In fact, because of the maintained presence of imperfect competition, one is inevitably faced with the second-best problem; removal of one set of distortions such as domestic tariffs may well reduce welfare. Furthermore because the economy faces less than perfectly elastic demand curves for its products, there can be changes in the terms of trade. It is, of course, possible in a partial-equilibrium framework with strong assumptions on tastes and technology to get definite predictions as to some of these effects. Yet the lack of generality of these theoretical results is yet another reason to resort to empirical analysis.

IV. Some Quantitative Results and Model Comparisons

In this section some illustrative results are given for various models using as a basis for comparison some trade liberalization experiments.[15] The models considered are (*i*) a strict neoclassical competitive constant returns model; (*ii*) the general equilibrium industrial organization model with product differentiation, and (*iii*) the general equilibrium industrial organization model without product differentiation.

In each case the models were scaled to a 1976 Canadian data set (the construction of which is described in my earlier study, Appendix A), together with extraneous econometric estimates on various elasticities. The model has 29 industries: 20 manufacturing industries treated as potentially noncompetitive in the industrial organization models and the other 9 as competitive at the world rental rate. In general the elasticity values used are consistent with those used in other applied trade models. There are 29 domestic industries with import elasticities that range from -1.0 to -3.0, and export elasticities that range from -0.9 to -3.1. Of the 29 industries, 20 are the two-digit SIC manufacturing industries treated as potentially noncompetitive. Scale elasticities of plant cost curves, defined as the ratio of marginal to average cost, within these industries range from 0.75 to 0.90. It is also the case, not surprisingly, that industries with high scale elasticities in the base equilibrium (low unexploited scale economies) are also relatively labor intensive. The 9 nonmanufacturing industries are treated as constant returns industries. The levels of protection, in the form of ad valorem tariff and equivalent nontariff barrier (NTB) rates, are in the normal range. The average domestic "tariff" on manufacturing products is 11 percent and the average foreign tariff on manufacturing products is 16 percent. These tariffs were those in place in 1976.

The industrial organization models were scaled relative to a short-run equilibrium with the number of firms held constant in each noncompetitive industry. Alternative long-run equilibria are simulated under various assumptions as to firm behavior and policy parameters. The algorithm used solves the short-run equilibrium using a conventional numerical interpolation procedure. The long-run equilibrium is solved by an iterative procedure which essentially mimics the Marshallian textbook procedure of exit and entry in response to profits and losses, together with an updating of perceived demand elasticities. The policy simulations reported involve a

[15] In our 1983 paper, Cox and I give detailed results, including sensitivity analysis, of the use of the model for trade liberalization policy simulations for Canada.

TABLE 2—TRADE LIBERALIZATION EXPERIMENTS IN ALTERNATIVE MODELS

	PC Model		*IO* Model with *PD*		*IO* Model without *PD*	
Variable	*UFT*	*MFT*	*UFT*	*MFT*	*UFT*	*MFT*
1. Aggregate Welfare Gain	0.0	2.4	2.7	6.2	4.1	8.6
2. Percent Change in Aggregate Labor Productivity	2.9	7.6	19.6	32.6	20.6	32.6
3. Percent Change in Index of Length of Production Runs	–	–	37.2	67.7	41.4	66.8
4. Percent Change in B-G-L Intra-Industry Trade Index	−5.2	−8.4	−0.6	−1.7	−0.7	−1.7
5. Interindustry Shifts in Labor Force	2.7	6.2	3.9	6.1	3.9	6.1

Notes: All changes are relative to base 1976 equilibrium. Lines 1: Welfare gain is Hicks equivalent variation as a percent of base *GNE*; 2: Labor productivity is weighted average of labor productivity in each sector with weights equal to industry shares in value of total output; 3: Index of production runs is weighted average of representative firms' production run length in each industry; 4: B-G-L Intra-industry trade index is trade weighted average of Balassa-Grubel-Lloyd index $\equiv 1-(X_i - M_i)/|X_i + M_i|$, where $X_i \equiv$ value of exports and M_i value of imports (in sector i); 5: Interindustry shift in labor force is percent of total labor force moving between industries. Columns: *PC* Model = Perfectly Competitive Model; *IO* Models = Industrial Organization Models; *PD* = Product Differentiation.

unilateral cut of all domestic tariffs (including *NTB* ad valorem equivalents), or unilateral free trade (*UFT*), and a multilateral cut of all tariffs (and *NTB* ad valorem equivalents), foreign and domestic (*MFT*).

Table 2 provides an indication of the type of aggregate results the different models yield. The most striking results are the welfare costs. The perfect competition (*PC*) model, given the removal of all domestic tariffs, or unilateral free trade, yields a net welfare gain of zero percent, while the industrial organization model without product differentiation yields a welfare gain of 4.1 percent of base national income. Under multilateral free trade the *PC* model yields a gain of 2.4 percent while the non-*PD*/*IO* model gives a gain of 8.6 percent. The estimated gains to free trade are quite significant using the *IO* model, and much larger than most studies using either partial-equilibrium triangle methods, or competitive general equilibrium simulation methods. The industrial organization model with product differentiation gives smaller welfare gains than without *PD*. The reason is that the rationalization effects of free trade lowers the number of domestic firms and hence the number of domestic and foreign products; this has an adverse effect on welfare given the *PD* assumption. It seems difficult to attach much significance to this result for two reasons. First, if the alternative hypothesis that any domestic good dropped will be replaced by a foreign substitute, the negative welfare effect will disappear. Second, the elasticities approach to product differentiation may be an inadequate description of consumers' taste for product diversity; thus quantitative estimates to industries dropping and adding products may be seriously biased.

The rationalization and procompetitive effects in the *IO* models are clearly responsible for the larger welfare gains. Aggregate labor productivity increases dramatically. Notice the large increases in average production runs—in the order of 50 percent under both *UFT* and *MFT*. Indeed the domestic tariff alone seems responsible for much of the increased scale economies at the industry level. A 50 percent increase in output at the plant level with a scale elasticity of 0.8 implies a fall in average cost of approximately 8 percent. These are, of course, just averages—

TABLE 3—RELATIVE INDUSTRY OUTPUT CHANGES BY MODEL AND POLICY EXPERIMENT

	UFT		*MFT*	
Industry	*IO* Model	*PC* Model	*IO* Model	*PC* Model
1. Food and Beverage	0.1261	−0.0821	0.2907	0.0077
2. Tobacco	0.2831	−0.0814	0.3178	−0.0869
3. Rubber and Plastic	0.3206	−0.0220	0.4276	0.0489
4. Leather	−0.1394	−0.3326	−0.1314	−0.3848
5. Textiles	0.1517	−0.1717	0.9417	0.5741
6. Knitting Mills	−0.2405	−0.5484	0.0674	−0.3945
7. Clothing	−0.2277	−0.6332	0.6842	−0.2093
8. Wood	0.1020	0.0247	0.1189	0.0461
9. Furniture and Fixtures	−0.0907	−0.2011	−0.1837	−0.2903
10. Paper and Allied Products	0.1847	0.0365	0.9578	0.9162
11. Printing and Publishing	0.0589	−0.0292	0.3423	0.1900
12. Primary Metals	0.4285	0.2686	0.3753	0.0709
13. Metal Fabricating	0.2094	0.0044	0.2261	−0.1863
14. Machinery	0.0541	0.0234	−0.0703	−0.1863
15. Transportation Equipment	1.0553	0.5448	1.2184	0.3727
16. Electrical Products	0.0580	−0.1083	0.0190	−0.1799
17. Non-Metallic Mineral Production	0.1947	0.0110	0.2521	0.0172
18. Petroleum and Coal	0.1182	0.0654	0.2596	0.1820
19. Chemical Products	0.1366	0.0103	0.2885	0.1265
20. Misc. Manufacturing	−0.0044	−0.2005	−0.1046	−0.2842
21. Agriculture	0.0479	−0.0482	0.6063	0.5210
22. Forestry	0.0955	0.0243	0.3123	0.3592
23. Fishing	0.0374	0.0426	0.3214	0.2472
24. Mining	0.0507	0.0393	0.2846	0.2683
25. Construction	0.0138	−0.0090	0.0404	0.0129
26. Transportation	0.0088	0.0021	0.0152	0.0249
27. Communication	0.0085	−0.0126	0.0237	0.0106
28. Electric, Power and Gas	0.0716	0.0098	0.1728	0.1033
29. Others	0.0296	−0.0098	0.0623	0.0309
Simple Correlation Coefficient	.0061		.0079	
Rank Correlation Coefficient	.3079		.1661	

the pattern across industries is highly variable with industries with strong scale economies rationalizing a great deal more than others. As most of the productivity gains translate into increases in the real wage, the labor-intensive industries do particularly badly under free trade. On the other hand, industries with significant scale economies are also relatively capital intensive, so they do relatively better on two counts.

The overall impact on trade patterns is quite different in the neoclassical than in the *IO* model. Both unilateral and multilateral free trade lead to significant decreases in interindustry trade in the neoclassical model. The *IO* models yield relatively little change in intra- vs interindustry trade. The intersectoral aggregate shift in labor is however comparable for both models, in the case of free trade, with slightly less labor shifting intersectorally in the neoclassical model for *UFT*. In aggregate terms the results indicate that for a trade liberalization experiment the *IO* model gives intersectoral resource shifts on a comparable level to the neoclassical model, but that changes in trade patterns are quite different between the two models.

At the industry-sectoral level the models yield dramatically different results. In Table 3 relative industry output changes under *UFT* and *MFT* for the two models are presented. The two models implemented on the same data base and elasticity estimates yield dramatically different conclusions. The simple correlation coefficients between relative outputs changes for both policy experiments is less than 0.01. The hypothesis of no correlation in levels or ranks is accepted at the 99

percent level of significance. To give another example, in the case of domestic tariff removal the two models give opposite signs on expansion vs. contraction of industry in 10 of the 29 cases.

In general terms both *UFT* and *MFT* result in resource transfers of the scarce factor (labor) to the manufacturing sector. For *UFT*, of course, this is contrary to the conventional perfectly competitive small open economy result in which case resources are transferred *out* of manufacturing. This conclusion, suggested long ago by Wonnacott-Wonnacott, results from the presence of a protected and inefficient manufacturing sector with unexploited scale economies. The procompetitive effects of the removal of domestic tariffs alone forces lower markups and the exit of domestic firms in protected industries. With lower average costs of production due to the greater output per firm, comparative advantage shifts toward manufacturing in response to the tariff cut and resources flow in that direction. It is evident the conventional model and the industrial organization model give significantly different views on the pattern of interindustry adjustment to trade.

The important difference between these models obviously hinges on the presence of scale economies at the level of the plant, and the manner in which prices are set. The models are similar in that both assume freedom of entry and exit to all industries. To give some idea of the significance of alternative pricing assumptions, in Table 4 results on welfare and productivity to *MFT* of varying the weights on the two basic pricing hypotheses are reported. Going down the column in Table 4 attaches greater weight to the collusive pricing hypothesis—this can be thought of as exogenously moving from an assumption that industry conduct is relatively (monopolistically) competitive to an assumption that industry conduct is relatively collusive. The table clearly demonstrates that assumptions as to the degree of competition, or nature of industry conduct, are crucial to determining the quantitative effects of trade liberalization. If industries are relatively collusive, the procompetitive effects of trade liberalization in a small open economy are much greater.

TABLE 4—*MFT* RESULTS AS PRICING HYPOTHESES VARIED

Weight Attached to Collusive Pricing Hypothesis[a]	*MFT* Welfare Gain	*MFT* Change in Aggregate Labor Productivity
0.2	4.31	22.26
0.4	6.87	28.44
0.5	8.59	32.62
0.6	10.75	37.99
0.8	16.32	56.98

[a]The results reported in Tables 2 and 3 correspond to a weight of 0.5 on the collusive pricing hypothesis.

Some general comments on the experience with the industrial organization models is offered although no results are reported in detail. Further results are available in David Cox's and my paper (1983), and my earlier study. First, the *IO* models are generally more parameter sensitive than competitive models. This is particularly true of parameters describing scale economies, the degree of collusive pricing, and basic import and export elasticities. This in turn means that it is extremely important to do sensitivity analysis and to obtain the most reliable estimates of parameter values available. The divergence between the competitive model and industrial organization model holds up for a wide variety of policy experiments. In one respect this is not a comforting result since it suggests that a large number of studies on resource allocation policies in small open economies may be seriously misspecified in both their general aggregate quantitative impact and the sectoral composition of the effects. On the other hand, the industrial organization models themselves are undoubtedly misspecified. Aggregation, if anything, is less plausible given imperfect competition and scale economies. Furthermore industries with significant entry barriers other than scale economies exist and these barriers are not adequately captured in the model. Multinational enterprises and foreign direct investment have yet to be incorporated adequately. My guess, however, is that failure to deal with these problems does not mean that the "true" results are more likely to be in conformity with the competitive model, but rather farther apart.

V. Conclusion

An applied general equilibrium model of a small open economy with an imperfectly competitive sector has been described and illustrative results on a Canadian 1976 data set given. The details of industrial organization in the noncompetitive sectors include economies of scale, explicit price-setting firms, and product differentiation. Some general equilibrium simulations on trade liberalization were carried out and compared with a competitive model implemented on the same data set. The major conclusion is that the estimated welfare gains from trade liberalization are substantial in the industrial organization model and on the order four times larger than the gains estimated from the competitive model. Furthermore the models differ significantly in their predictions as to the interindustry pattern of adjustment to trade liberalization. Intra-industry adjustment is an important avenue for resource reallocation in the industrial organization models. With trade liberalization, firms achieve larger scale economies in response to larger world markets and competitive pressure from abroad.

The lessons learned from this analysis should be useful in the examination of policies impinging on resource allocation in small open economies. It is both feasible and useful to incorporate industrial organization features in applied general equilibrium models. Furthermore, the failure to do so may lead to quite different conclusions in the empirical evaluation of policy at both the industry and aggregate level.

REFERENCES

Alchian, Armen, "Costs and Outputs," in M. Abramovitz, ed., *The Allocation of Economic Resources*, Stanford: Stanford University Press, 1959.

Balassa, Bela, *Trade Liberalization Among Industrial Countries*, New York: McGraw-Hill, 1966.

______, *European Economic Integration*, Amsterdam: North-Holland, 1975.

Boadway, Robin and Treddenick, John, "A General Equilibrium Computation of the Effects of the Canadian Tariff Structure," *Canadian Journal of Economics*, August 1978, *11*, 424–46.

Brander, James, "Intra-Industry Trade in Identical Commodities," *Journal of International Economics*, February 1981, *11*, 1–14.

Brown, Fred and Whalley, John, "General Equilibrium Evaluations of Tariff-Cutting Proposals in the Tokyo Round and Comparisons to More Extensive Liberalization of World Trade," *Economic Journal*, December 1980, *90*, 838–66.

Cline et al., W. R., *Trade Negotiations in the Kennedy Round: A Quantitative Assessment*, Washington: The Brookings Institution, 1978.

Corden, W. M., "Economies of Scale and Customs Union Theory," *Journal of Political Economy*, April 1972, *80*, 465–75.

______, *Trade Policy and Economic Welfare*, Oxford: Oxford University Press, 1974.

______, "The Costs and Consequences of Protection: A Survey of Empirical Work," in P. Kennen, ed., *International Trade and Finance*, London: Cambridge University Press, 1976.

Cox, David and Harris, R., "Trade Liberalization and Industrial Organization: Some Estimates for Canada," Discussion Paper No. 523, Department of Economics, Queen's University, 1983.

Deardorff, A. V. and Stern, R. M., "A Disaggregated Model of World Production and Trade: An Estimate of the Impact of the Tokyo Round," *Journal of Policy Modelling*, 1981, *3*, 127–52.

Dixon et al., P., *ORANI: A Multi-Sectoral Model of the Australian Economy*, Amsterdam: North-Holland, 1981.

Eastman, H. C. and Stykolt, S., *The Tariff and Competition in Canada*, Toronto: University of Toronto Press, 1966.

Evans, H. D., *A General Equilibrium Analysis of Protection*, Amsterdam: North-Holland, 1972.

Grubel, H. G. and Lloyd, P. J., *Intra-Industry Trade: The Theory and Measurement of International Trade in Differentiated Products*, New York: Wiley & Sons, 1975.

Harris, R. G., *Trade, Industrial Policy and*

Canadian Manufacturing, Toronto: Ontario Economic Council, 1984.

Helpman, Elhanan, "International Trade in Presence of Product Differentiation, Economies of Scale and Monopolistic Competition: A Chamberlin-Hecksher-Ohlin Approach," *Journal of International Economics*, August 1981, *11*, 305–40.

Ingram, W. D. and Pearson, S. R., "Economies of Scale, Domestic Divergences, and Potential Gains from Economic Integration in Ghana and the Ivory Coast," *Journal of Political Economy*, October 1980, *88*, 994–1008.

Krugman, Paul R., "Increasing Returns, Monopolistic Competition, and International Trade," *Journal of International Economics*, November 1980, *9*, 469–79.

Lancaster, Kelvin, "Intra-industry Trade under Perfect Monopolistic Competition," *Journal of International Economics*, April 1980, *10*, 151–75.

Lloyd, P. J., "Protection Policy," in F. H. Gruen, ed., *Surveys of Australian Economics*, Sydney: Allen and Unwin, 1978.

Magee, Steven P., "The Welfare Effects of Restrictions on U.S. Trade," *Brookings Papers on Economic Activity*, 3:1972, 645–707.

Negishi, Takashi, "Monopolistic Competition and General Equilibrium," *Review of Economic Studies*, June 1961, *28*, 196–201.

Panzar, John C. and Willig, Robert D., "Economies of Scope," *American Economic Review Proceedings*, May 1981, *71*, 268–72.

Shoven, John B. and Whalley, John, "Applied General Equilibrium Models of Taxation and International Trade," National Bureau of Economic Research, Stanford, 1983.

Spence, A. Michael, "Product Selections, Fixed Costs and Monopolistic Competition," *Review of Economic Studies*, June 1976, *43*, 217–53.

Williams, J., *The Canadian-U.S. Tariff and Canadian Industry: A Multisectoral Analysis*, Toronto: University of Toronto Press, 1976.

Wonnacott, R. J. and Wonnacott, P., *Free Trade Between the United States and Canada: The Potential Economic Effects*, Cambridge: Harvard University Press, 1967.

Economic Council of Canada, *Looking Outward, A New Trade Strategy for Canada*, Ottawa: Information Canada, 1975.

Quasi-Pareto Social Improvements

By YEW-KWANG NG*

The Pareto criterion is widely accepted as a sufficient condition for an improvement in social welfare. If someone is made better off and no one worse off, there seem to be no acceptable grounds to reject the change. However, most, if not all, changes in the real world involve making some better off and some (no matter how small the number) worse off. Thus the Pareto criterion in itself is of little practical use. On the other hand, the search for a widely acceptable criterion for (social) welfare improvements beyond Pareto, around the 1940's, in the form of compensation tests and the like, seems to have encountered overwhelming objections. As a result, little if any advance has since been made. We still have not gone beyond Pareto as far as a widely acceptable welfare criterion is concerned.[1]

In Section I, I propose a widely acceptable welfare criterion beyond the Pareto principle. The proposal consists in amending the well-known Kaldor-Hicks-Scitovsky double compensation test by requiring it to be satisfied for each and every (usually income) group of individuals. This amendment rids it of its main objection on the ground of distributional considerations (for example, making the poor poorer and the rich much richer is usually not regarded as a good change).[2] It may be thought that this amendment is too restrictive and makes the criterion virtually useless. Section II illustrates the wide applicability of the criterion by using it to justify a specific proposal for improving the allocation of water. Imperfect knowledge, administrative costs, and the diversity of individual preferences make it impossible, in most cases, to design a change or a policy that makes every individual better off. But it may be possible to design one that makes every group better off, satisfying the criterion. Second, since the criterion is meant as a sufficient, not as a necessary, condition for a social improvement, its acceptance does not prevent one from going beyond the criterion to accept, say, changes that make the poor better off and the rich worse off. Third, when combined with the third-best equality-incentive argument, this criterion leads to a much more forceful principle of a dollar is a dollar *irrespective* of income groups (Section III). This provides a powerful simplification in economic assessment (of any change, policy, etc.) in general, and in cost-benefit analysis in particular. It provides a formal justification for the separation of equity and efficiency considerations (see Richard Musgrave, 1969; A. C. Harberger, 1971, among others) despite the presence of second-best factors and other complications (Section IV).

I. A Proposed Welfare Criterion of Group-Specific Compensation Tests

It is the hypothetical nature of compensation (i.e., without having to execute the actual compensation) that makes compensation tests interesting. With actual compensation to make everyone better off, the change becomes a Pareto improvement and no separate welfare criteria are necessary. But the hypothetical nature also attracts two separate objections. First, both Nicholas Kaldor's criterion (1939; ability of gainers to compensate losers) and John Hicks' criterion (1940; inability of losers to bribe gainers to oppose the change) may be logically inconsistent, since both a change and its reverse

*Reader in Economics, Monash University, Clayton, Victoria, Australia 3168. I am grateful to my colleague Ross Parish for stimulating discussion and to a referee for helpful suggestions. I have also benefited from seminar discussions of earlier versions of Section III at Monash (1976), Virginia Polytechnic Institute, New York, and Yale universities (1978).

[1]For surveys of the debate on welfare criteria, see Ezra Mishan (1969), John Chipman and James Moore (1978), and my book (1979; 1983, ch. 3).

[2]Another objection of possible inconsistency after repeated applications is discussed in the Appendix, part A, and regarded as insignificant in practice and hence not a compelling objection to a *practical* criterion.

(back to the original position) may satisfy the same compensation test. Second, hypothetical compensation need not ensure a social improvement as the smaller amount (in monetary terms) of loss by the losers may be socially more important than the gain to the gainers for some reason. Thus, if the poor lose and the rich gain, the change may not be regarded as an improvement even if it satisfies all compensation tests. Hicks attempts to overcome this second problem by arguing that, with repeated application of compensation tests, "there would be a strong probability that almost all...[individuals] would be better off after the lapse of a sufficient length of time" (1941, p. 111). (See also Harold Hotelling, 1938; James Buchanan and Gordon Tullock, 1962, pp. 77–80; Harvey Leibenstein, 1965.) While this is certainly a very fruitful way of strengthening the compensation principle (as developed further by A. Mitchell Polinsky, 1972), it does not eliminate the problem completely as changes that persistently hurt some particular group cannot be ruled out completely, intertemporal substitution is not perfect, and the aged cannot live much longer.

To overcome the two above problems of compensation tests, I propose the following two amendments. The first amendment was made by Tibor Scitovsky (1941). Effectively, a change must satisfy both the Kaldor and the Hicks criteria. Apart from exceptional cases (see my book 1979; 1983, Appendix 4A), little attention has to be paid to this amendment. Especially for changes whose effects are thinly spread across a large number of individuals, the differences between ΣCV (the sum of compensating variations in income across individuals) and ΣEV (sum of equivalent variations) are likely to be small in comparison to the inaccuracies arising from difficulties in data collection. Thus if we are certain that one of the two compensation tests is satisfied, the other is also most certainly satisfied.[3]

The second amendment is concerned with the distributional issue. I. M. D. Little's (1950; 1957) approach was to impose the requirement that any distributional effect must be favorable. This does not free Little's criterion from the charge of logical inconsistency, as it requires only (at least) one of the two compensation tests to be satisfied. I have offered a defense of this aspect of the Little criterion elsewhere (see my book, pp. 68–72; see also Kotaro Suzumura, 1980). Even accepting this defense, the question of how to decide whether the distributional effects are "favorable" is still left unsolved. To make our welfare criterion widely acceptable, the following amendment is proposed.

Instead of requiring the satisfaction of compensation tests across all individuals, let us require the same satisfaction within each income group (or other group of the same "deservingness," if income is not the only problem). How finely income groups should be defined is a matter of choice. On one extreme, if all individuals affected by the change fall within the same income group, our criterion is equivalent to the Kaldor-Hicks-Scitovsky criterion. On the other extreme, if an income group is defined narrowly enough such that each individual income-earner is a distinct income "group," our criterion collapses into the Pareto criterion. This latter extreme is unlikely to be relevant in practice for changes that affect a large number of individuals. It is practically impossible to distinguish an income-earner of \$23,451 per annum and another of \$23,452 per annum. For most practical purposes, the following classification is sufficient: the destitute, the very poor, the poor, the just-below average, the average, the just-above average, the rather well-to-do (usually called

[3]Due to the fact that CV and EV are measured at given prices while compensation tests are based on feasible compensation which may involve changes in prices, a positive ΣCV (ΣEV) is not necessarily equivalent to the satisfaction of the Kaldor (Hicks) criterion, as demonstrated by Robin Boadway (1974). Quoting from my earlier book, "Nevertheless in the real economy where a certain change is small relative to GNP, the payment of compensation is unlikely to change prices significantly. Even if prices are changed, the effects of the changes are likely to be negligible compared with inaccuracies in data collection. Thus, if ΣCV is big enough to overbalance data inaccuracies, we can be quite safe in concluding that full compensation is possible" (p. 98). See also J. S. Dodgson (1977).

the middle class?), the very well-to-do (upper middle?) the rich, the very rich, and the extremely rich. Of course, a finer classification may be used with actual income ranges as well as supplementary conditions (number of dependents, etc.) specified if desired.

The social marginal utility of a dollar is taken as approximately the same across all income-earners within a given income group. Thus, if compensation is possible within each income class, social welfare may be taken as increased. (For complications arising from a shift in income status from one group into another, see the Appendix, part B). Some people may be willing to go further by allowing net losses in higher income groups provided they are compensated by net gains in lower income groups. I do not propose to strengthen our welfare criterion in this way because: 1) the criterion is meant as a sufficient condition for a social improvement, not a necessary condition; 2) I want to command as wide an acceptance as possible; 3) I have a different method of dealing with the problem of equality (Section III below).

It may be thought that, by requiring the satisfaction of the compensation test for each income group, our criterion is very demanding if a fine grouping is adopted such that very few changes can satisfy our criterion. However, this is not the case, rather surprisingly. The next section illustrates how the criterion may be used to sanction certain changes. Section III combines the use of our criterion with the third-best equality-incentive argument to arrive at the powerful conclusion of "a dollar is a dollar" (across all income groups).

II. Water Pricing: An Illustrative Application of Our Welfare Criterion

Water restrictions were introduced in Melbourne in 1982 and continued until now (end of 1983) in a bid to preserve the water supply. It is natural for an economist to ask: why is the price system not used to allocate scarce water? That water is essential for life is not the answer as many other things (food, shelter, clothing, etc.) are also essential. Moreover, a minimum amount (less than 5 percent of present consumption) essential for life can be allowed free and the excess charged according to costs. The cost of charging for water is also not the answer. The "dominant opinion in the field of municipal water supply seems to be that universal metering produces gains that are worth the cost" (Jack Hirshleifer, J. C. de Haven, and J. W. Millman, 1969, p. 45). Moreover, Melbourne *already* has a metering system with annual reading to charge for *possible* excess water consumption. The free amount of water consumption (free entitlement) is proportional to the fixed charge which is in turn proportional to the estimated property value. But these free entitlements are such that most households do not have to pay any excess water bill even when consuming as much water as desired. Thus, the bulk costs (metering) of charging for water are incurred without gaining the bulk benefit (incentive to conserve water in accordance to its marginal cost by the majority of consumers)—a most uneconomical situation.

My suggestion for changing the present system into a full pricing system (drastically lowering the free entitlements) was referred by the Water Supply Minister to the Melbourne and Metropolitan Board of Works (MMBW) for consideration. After a discussion with the Director of Finance of MMBW, I understand that one of the most important factors they are concerned with is the implication of the proposal on the cross subsidization in the present system. This cross subsidization exists because consumers of high property values are paying high fixed charges. Though they are entitled to proportionately more free water, they seldom use the maximum free entitlements. The situation is more or less as depicted in Figure 1. On average, consumers with property values higher than $\bar{P}$ consume less than their free entitlements. Those consuming much less than their free entitlements are thus effectively subsidizing those consuming more or only slightly less than their free entitlements (excess water consumption is charged at the same rate as used in the calculation of the free entitlement allowed by the fixed charge).

If the free entitlement and fixed charges are reduced proportionately, those presently

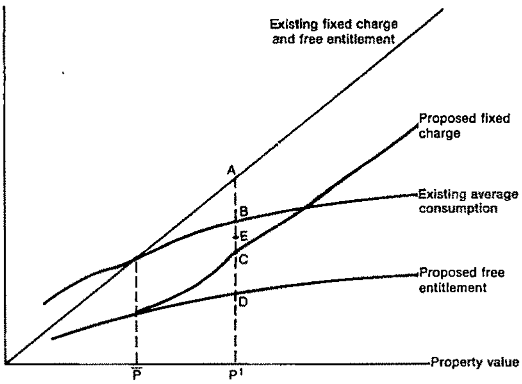


FIGURE 1

consuming much less than their free entitlements (mainly those with high property values) will gain. Moreover, the per unit price for excess water may have to increase to compensate for the loss of fixed charges. Thus those presently consuming more than their free entitlements (mainly those with low property values) may lose. (Only "may," not "will," since all consumers gain from the more efficient system and freedom from arbitrary water restrictions.) Thus, unless the present cross subsidization is regarded as undesirable in some sense, the change is not necessarily desirable despite the pure efficiency gain. However, we may devise a system to capture the efficiency gain without (on average) changing the cross subsidization. This can be done by reducing the fixed charges by proportionately less than the reduction in free entitlements for consumers with high property values.

As illustrated in Figure 1, for example, the free entitlements can be reduced to a fraction of the existing average consumption and the fixed charges are adjusted accordingly such that any consumer who consumes the old *average* consumption at his property value will pay the same total amount (fixed charge plus excess water bill) as in the present system. For example, a consumer whose property is valued at P^1 is paying under the present system a fixed charge of P^1A which also measures his free entitlement. If he consumes the average amount of water P^1B, he pays no excess water bill, as is typically the case. In the proposed system, his free entitlement is reduced to P^1D. If he still consumes the average amount P^1B, he has to pay BD for excess water consumption (the unit price is kept unchanged). But his fixed charge is reduced by the same amount AC ($= BD$ by construction). Hence he will be no better off and no worse off if he continues to consume the same amount P^1B. However, he is now given the opportunity to reduce his water bill by consuming less water. Thus, for the average consumer (those who consume the average amount at their property values), no consumer is made worse off and most consumers are made better off by having the opportunity to reduce their water bill. The total revenue to MMBW would be unchanged if consumers did not take up this opportunity to save by conserving water. If they did save, the amount of water conserved would be worth the reduced revenue collected, assuming that the unit price is originally fixed at an appropriate level.

Even ignoring problems of costs of changing to the new system, possible difficulties associated with making the cross subsidization more transparent, etc., the above proposed change does not satisfy the Pareto criterion. This is so because not all consumers are average consumers. For example, consumers with properties valued at P^1 may consume more or less than P^1B (the average figure for these consumers). With the new system, those consuming less than the average will gain (on top of the opportunity to save water) since their fixed charges will be reduced by the amount AC and their excess water bills, even if they continue to consume the same amounts of water, will be only, say, DE. Conversely, those consuming more than the average will lose (ignoring the gain from the opportunity to save water). Thus, some consumers will be made better off and some may be made worse off. The Pareto criterion is insufficient to sanction the proposed change. However, at each property value, ignoring the efficiency gain, the loss of some consumers is exactly offset by the gain of others. We may thus expect that our double compensation test will be satisfied for each group of consumers (one at each property

value) when the efficiency gain is taken into account.[4] Since the proposed change is unlikely to affect any consumer so severely as to change his status in the income grouping by more than one step (for complications due to such changes, see the Appendix, part B), and since consumers of the same property value may be taken as approximately similar in terms of the rich-poor scale, our welfare criterion is satisfied by the proposed change discussed above.

The principle illustrated in the application of this welfare criterion has wide applicability. Most changes affect different individuals differently—some gain, some lose. Even if a change is carefully designed so as to leave some net gain to every group, some nonaverage individuals may still lose. The existence of administrative costs, the lack of perfect information, and/or the principle of equal treatment for equals (consumers of the same property value in the example above) preclude in most cases the possibility of designing a change that will make every individual better off. The use of our welfare criterion thus provides an acceptable middle ground between the Pareto criterion that is almost never satisfied and the ordinary compensation principle that may make a whole group of individuals (for example, the poor) significantly worse off.

III. A Dollar is a Dollar Irrespective of Income Group

Ignoring the differences between EV and CV as insignificant and/or largely offsetting for most changes (at least those with their effects thinly spread; for a way to handle exceptions, see my book, Appendix 4A), and ignoring the effects on relative prices as again insignificant and/or largely offsetting (see my book, p. 98), we may regard the criteria of ΣCV, ΣEV, and the Kaldor and Hicks compensation tests as approximately equivalent. These criteria treat a dollar (gain or loss) as equal to another dollar, to whomsoever it goes, the rich or the poor. The main objection to such a principle of "a dollar is a dollar" is the question of inequality in income distribution since many people regard a dollar to the poor as satisfying more important needs than a dollar to the rich, and believe that the relevant benefits or costs should thus be valued accordingly in the application of a welfare criterion or in cost-benefit analysis. In particular, differential income (or distributional) weighting and other preferential treatment between the rich and the poor, as well as nonmarket allocation (for example, time limits on metered parking irrespective of willingness to pay) are regarded as desirable despite their efficiency loss. These will be referred to as purely equality-oriented preferential policies. Using our welfare criterion of group-specific compensation tests, I show in this section that the objective of income equality can be better achieved through income taxation. By thus supporting the principle of a dollar is a dollar, I make, somewhat paradoxically, my own welfare criterion redundant. In other words, after justifying a dollar is a dollar, the group-specific proviso in compensation tests need no longer be insisted on. (This paradox is explained in Section V.)

Essentially, the objective of achieving a more equal distribution of income is better achieved through income taxation even if disincentive effects are involved since purely equality-oriented preferential policies have efficiency costs[5] as well as disincentive effects. A rational individual without money illusion will not only take into account the amount of income earned, but also what he will get from the income. If having a higher income does not enable one to buy more parking space but rather means that one has to pay more for the same thing, and/or means that

[4] Since the change is unlikely to affect prices by a significant extent, we may expect that $\Sigma CV > 0$, $\Sigma EV > 0$, and that both the Kaldor and Hicks compensation tests will be satisfied for each group of consumers, assuming no noneconomic objection to the new system as such.

[5] On the potential enormous efficiency costs of the application of distributional weights, see Harberger (1978).

projects in his favor are less likely to be undertaken, then the extra income will be worth less than it otherwise would be. It is as though the income tax rate has increased. The same degree of disincentive will apply in both cases unless income is valued not so much for its purchasing power but mainly as a status symbol. However, among those people for whom status considerations are important, it is more likely that the pre-tax rather than the post-tax income will be used as an indication. Hence the same degree of incentive still applies.

In an ideal first-best world where costless and neutral lump sum transfers (fixed according to potential rather than actual income) are feasible, it can easily be seen that a dollar is a dollar. A dollar must be treated as a dollar to maximize efficiency, and any desired level of equality can be achieved by lump sum transfers. But the real world is not first best as lump sum taxes on potential incomes are not feasible. Actual redistributive policies may take the form of (*i*) measures that improve both efficiency and equality such as the dismantling of artificial barriers to the equality of opportunity, (*ii*) progressive income taxation, and (*iii*) purely equality-oriented preferential policies. Ideally, of course, all of type (*i*) measures should be adopted. But usually, after all feasible measures of that type have been used, equality is still not regarded as sufficiently attained. Hence, both types (*ii*) and (*iii*) are also used. The inferiority of the purely equality-oriented preferential policies may be gauged by the following proposition.

PROPOSITION 1: *For any alternative (designated A) using a system (designated a) of purely equality-oriented preferential treatment between the rich and the poor, there exists another alternative, B, which does not use preferential treatment, that makes no one worse off, achieves the same degree of equality (of real income, or utility) and raises more government revenue, which could be used to make everyone better off.*

Note that this proposition is applicable even in the case where the preferential treatment just happens to be consistent with second-best efficiency considerations. Because, alternative B, which does not incorporate purely equality-oriented preferential policies, but, instead, uses system b, designed for efficiency purposes only, would already, within that system b, incorporate the second-best efficiency considerations with which system a just happens to be consistent. In other words, the principle of a dollar is a dollar does not preclude adjustments based on some efficiency considerations such as externalities, second best, etc. But it excludes the purely equality-oriented policies used in practice.

However, the *existence* of alternative B does not necessarily mean that it can be identified and implemented. If system a is *designed* to take account of second-best considerations, then system b can also be so designed. But system a may only be consistent with second-best considerations by *chance* rather than by design. In addition, the informational costs of designing a system consistent with the second-best considerations may be prohibitive.[6] Then we may not be able to identify system b. Thus, while alternative B may exist, it may not be feasible to implement. Thus, if we wish to strengthen Proposition 1 to be one about the existence of a *feasible* superior alternative B, it would apply only in a probabilistic sense. That it (the strengthened proposition) still applies in a probabilistic sense is due to the theory of third best. Just as it may be consistent with second-best considerations, system a may also be opposite to the requirement of second best. The theory of third best (see my 1977 article) can then be used to show that the expected gain is negative. Hence, as far as the second-best consideration is concerned, the use of system a involves negative expected gains. For simplic-

[6] Second-best taxation-pricing rules are typically very complicated, even if only the efficiency consideration is taken into account. For the current literature on optimal taxation, see James Mirrlees (1976, 1981); Agnar Sandmo (1976). Conditions making second-best considerations ineffective are rather stringent, for example, separability in the utility function (see Anthony Atkinson and Joseph Stiglitz, 1976; compare Theodore Bergstrom and Richard Cornes, 1983).

ity, we may thus assume that system *a* is neutral with respect to second-best considerations. (This, in fact, gives it an advantage.)

Proposition 1 is established in the Appendix, part C, under the usual assumption that individuals have identical preferences but different abilities. Since preferences do differ in practice, it may thus not be actually possible to make a Pareto improvement upon a system using purely equality-oriented preferential policies. We would then have to work with average individuals and construct an alternative that makes these average individuals (one at each income level) better off. Some nonaverage individuals would be made better off and some worse off. In principle, compensation may be effected to make no one worse off. But this is not practicable. However, while not all individuals can be made better off in practice, each income group could be made better off in the sense that overcompensation is hypothetically possible at each income group. Thus, while the Pareto criterion is not satisfied by the change, my welfare criterion of group-specific compensation test is. Thus this welfare criterion, when combined with the third-best equality-incentive argument, leads to the conclusion of a dollar is a dollar.

In itself, my argument does not preclude the joint treatment of equity and efficiency issues due to the second-best consideration (for example, see Dieter Bös, 1984). But such a sophisticated optimization procedure usually involves prohibitive informational and administrative costs (see my 1977 article), and is also probably politically infeasible as it may involve prices, taxes, etc. favoring the rich and *against* the poor in some (probably half of all) sectors. As far as I know, all preferential policies in the real world are purely equality oriented. In combination with this practical consideration, my argument also justifies the complete separation of equity and efficiency considerations.

Let us illustrate the argument with a utility possibility map. In Figure 2, U_P and U_R represent the levels of utility of two (groups of) individuals in the society. Starting from the initial position D on the utility possibility curve I, consider a project (or any change) that involves a negative aggregate net benefit

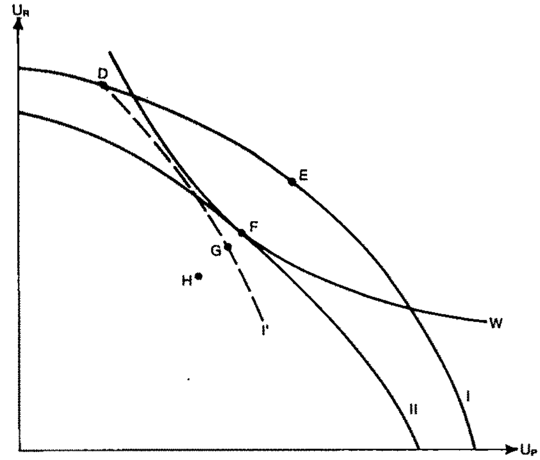


FIGURE 2

(the possibility curve moves inward to II), and a more equal distribution at point F. If the welfare contour[7] W through F passes above D, it appears that the change is desirable. Income weighting in cost-benefit analysis that sanctions such changes seems justified. However, to achieve the degree of equality represented by point F, it would be better to do so by income taxation, travelling along curve I from D to E. It is true that income taxation has disincentive effects and thus the point E is not sustainable. The disincentive effect will lead to a contraction of the utility possibility curve I inwards (not drawn) to pass through, say, point G. In other words, we cannot in fact travel along the utility possibility curve I, but have to travel along the utility feasibility curve I' in redistribution through income taxation. As drawn, G is inferior to F. But what is not commonly recognized is that point F is also not sustainable. If redistribution through income taxation from D to E will lead to the

[7]Murray Kemp and I (1976, 1977, 1982) show that a Bergson-Samuelson social welfare function (SWF) that is Paretian and based only on individual ordinal preferences does not exist for any given fixed set of individual preferences drawn from a wide domain. But we need not confine ourselves to ordinal preferences. In fact, I have shown elsewhere (1975) that a SWF must be an unweighted sum of individual cardinal utilities, if certain reasonable premises are accepted.

contraction of the utility possibility curve, so will redistribution through cost-benefit weighting from D to F. Abstracting from second-best considerations, II will contract by a roughly similar extent as I does. Hence, instead of point F, we will end up with H which is inferior to G. In the presence of second-best factors, we are uncertain as to which point we will end up with, but the expected average is somewhat below point H.

IV. Some Complications

In this section, I discuss factors that may render my general argument for a dollar is a dollar in Section III inapplicable. I show that while these considerations qualify my conclusion, the main thrust of the central argument is not much affected.

A. *Political Constraints on Redistribution through Taxation*

It is argued above that, instead of using weighting, quotas, or other kinds of preferential treatment to achieve the objective of equality (not as second-best correctives), it is better to adopt a more progressive income tax schedule. What if the taxation system cannot be changed to the desired structure due to political constraints? If it is true that we can't change the taxation system but we can effect redistribution by other means, my conclusion may have to be qualified accordingly, though there is still the ethical question of the desirability of doing good by stealth. Yet, why should the political constraint act only to prevent redistribution though taxation and not redistribution by other means? Maybe because the voters are rather irrational. I suspect, however, that on this issue, voters are very rational and practical. The upper and middle classes will not only vote a government out of office for carrying out drastic changes in taxation but also for carrying out other drastic redistributive measures. Especially in the long run, the forces that operate to prevent redistribution through taxation will also operate to prevent redistribution by other means. If we are thinking in terms of a distributional equilibrium, the distribution should be considered not in terms of money income but in terms of real income. Naturally, if the rich are penalized in other ways they will have less tolerance as regards the progressiveness of taxation. For example, had Australia been operating closer to the principle of a dollar is a dollar in the past, the reduction in the progressiveness of its income tax schedule undertaken in 1978 would probably not have been required. Both the rich and the poor would probably have been better off.

It is true that actual political decisions are affected by a host of factors and not just by an impartial consideration of a balance between equality and efficiency. However, equality and efficiency are important considerations and the fact that preferential treatment is an inefficient tool to achieve equality and efficiency has to be pointed out.

B. *Transaction Costs*

My analysis has been based on the assumption that the additional transaction costs associated with redistribution through more progressive income taxation are not higher than the transaction costs of its alternatives. Apart from the disincentive effect (which has been taken into account and is not subsumed under transaction costs), the costs associated with income taxation seem to fall mainly under 1) costs of administration on the parts of both the taxpayers and the collectors, 2) costs involved in tax evasion and enforcement, and 3) costs of tax lobby activities and the like. It is recognized that all of these forms of costs are substantial. But the relevant amounts are not total costs, only marginal costs. For good or for bad, income taxation will be with us in the foreseeable future. The incremental costs of administration of a more progressive tax system seem trivial. The costs of a change from one system to another may not be trivial, but will probably not be substantial. However, at least in the long run, the relevant comparison is the cost of administering two alternative systems, the difference between which is probably quite negligible. A more progressive tax system may however involve higher costs in encouraging more evasion and more

lobby activities. But the increase in progressiveness is in lieu of some system of preferential treatment which itself is a subject of evasion and lobbying. While this is a subject where a precise conclusion can hardly be expected, it does not seem probable that costs involved in the latter (i.e., preferential treatment) will be much lower than costs involved in the former (i.e., a more progressive tax system). On the other hand, the costs of administering a pure taxation system are almost certainly significantly lower than those of administering a system of taxation combined with preferential treatment in government expenditure. Hence, consideration of transaction costs seems to strengthen, not weaken, my central conclusion.

There may be specific cases whereby the use of an apparently "preferential" policy may be superior to making the income tax more progressive because of significantly lower transaction (evasion, lobbying, and administrative) costs of the former. But such a policy is justified on its efficiency consideration of lower transaction costs (relative to income taxation) and cannot be justified on the purely equality consideration advanced by egalitarian lobbyists.

C. *Ignorance of Benefit Distribution*

My argument is based on the assumption that individuals know the distribution of costs and benefits in government expenditure across income groups, or the details of preferential treatment, so that the incentives are the same as an equivalent pure income taxation system. In practice, this knowledge is unlikely to be perfect. On the other hand, most individuals do know the scale of income taxation. Does this asymmetrical knowledge mean that the disincentive effects of income taxation are more severe than an equivalent preferential expenditure system, as, in effect, argued by Martin Feldstein (1974, p. 152)?

In the absence of perfect knowledge, an individual has to base his choice on his estimates. From the fraction of knowledge he possesses, it seems that he is as likely to overestimate as to underestimate the degree of progressiveness implied in a given preferential expenditure system, depending on the psychology of the individual in question. Hence, on the whole, the degree of incentives is likely to be similar between the preferential expenditure system and the pure income taxation system.

D. *Redistributive Effects of the Project Itself*

I have argued that a dollar should be treated as a dollar irrespective of whether it accrues to the poor or to the rich. But this argument does not show that a billion dollars is always equal to a billion dollars. This point can be seen clearly by considering a simple example. Consider two alternative projects: project M will increase the incomes (after allowing for cost share) of one million individuals by \$10 thousand each, and project N will increase the income of one single random individual by \$10 billion. Ruling out costless lump sum transfers (that would make us indifferent between the two projects), it is clear that project M will be preferred to project N by all social welfare functions egalitarian in incomes. This preference is not based on valuing a marginal dollar to the rich as lower than a marginal dollar to the poor. Rather, it is based on treating the first dollar as more valuable than the 10 billionth dollar, whomever they go to. (Due to diminishing marginal utility or risk aversion, the *same* person typically regards the loss of \$10 thousand as more significant in utility terms than the gain of \$10 thousand, and the gain of \$10 billion as less than a million times the gain of \$10 thousand.) Hence, the equality-incentive argument I use above does not apply here. However, the equality-incentive argument can be used to dispel the possible belief that, since a project that itself creates inequality is inferior to one with the same aggregate net benefits but which does not create inequality, a project that creates equality must be preferable to one with the same aggregate net benefit but distributionally neutral. Consider a third project O that will yield the same aggregate net benefits of \$10 billion but be distributed across the economy in such a way that the poor will have much higher benefits and the rich have

negative benefits. While this may seem to be a good thing in itself, the incentive argument will show that project O is in fact inferior to project M. If project O were to happen as a natural event, it would be preferable. But if it is *chosen* instead of project M, it will produce disincentive effects.

From the above, it may be said that, for projects whose redistributive effects are marginal, one can simply choose in terms of aggregate net benefits; for projects whose redistributive effects are significant, we should prefer the one with less redistributive effects, given the same aggregate net benefits. This seems to lend support to the concept of a conservative social welfare function discussed by W. M. Corden (1974, p. 107).

E. *Preference for Working*

If an individual prefers to have his income by earning it instead of receiving it as a transfer welfare payment, then a cost-benefit analysis that does not take this preference into account may be misleading. This has been emphasized by M. L. Skolnik (1970). This kind of complication can be taken care of by appropriate shadow pricing. For example, in the particular case considered here, the main difference is the possible preference of an individual for earning his income instead of receiving some kind of dole money. This can largely be taken care of by putting an appropriate shadow price on employment. For a single person without dependents, his income from a low-paid job is likely to be sufficient to preclude him from receiving a subsidy. All that is needed is a low or zero income tax, so he will not have to suffer the feeling of being on the dole. For families with dependents, the subsidy can be effected in the form of, say, substantial child-endowment payments differentiated according to income levels. A fixed child-endowment is used in Australia with no one feeling ashamed of receiving it and the introduction of differentiation is unlikely to change this substantially.

F. *Unexpected Emergencies*

In times of unexpected emergencies such as earthquakes, wars, etc., certain necessities may be in very short supply. In principle, we could impose appropriately higher taxes on the rich and those who happen to own the goods in short supply and pay subsidies to the poor and the victims of the disaster. Then the policy of a dollar is a dollar could still be best. However, due to time lags, imperfect information, and the like, it may be practically infeasible to effect the required changes in taxes/subsidies in time. Rationing of basic necessities such as medical supplies (which also involves external economies) may then be the best practical solution. However, the possible desirability of violating the principle of a dollar is a dollar in such emergencies does not mean that the same is true for normal times.

G. *Nonincome Indices for Preferential Treatment*

My analysis concentrates on the use of income as the index for preferential treatment and redistributive taxation. But surely income is an imperfect measure of "deservingness," and nonincome variables such as health and age status are likely to enter distributional objectives. In particular, the use of age as an index for preferential treatment will create few, if any, disincentive effects, since one cannot change one's age. However, we can similarly use age as an index for the purpose of tax-subsidy. The purpose of giving assistance to the aged, for example, can be achieved without the additional efficiency costs of, say, giving free milk as some may not wish to drink milk. (Subsidized milk to schoolchildren may, however, be justified on the efficiency ground of merit wants; on merit wants as a possible efficiency ground, see my book, Section 10A.3). The consideration of nonincome factors does suggest that a single tax based on incomes only may not be sufficient; the tax-subsidy system may have to take nonincome factors into account.[8]

[8]For example, consider the argument of William Baumol and Dietrich Fischer (1979) that the use of discrimination in wage rate is much more efficient (less output foregone) than nondiscriminatory taxation in achieving equality. This is based on the detailed knowl-

A related problem of using income as the basis for taxation is that measured income may be a poor indicator of actual earning potential due to savings, risk bearing, etc. Thus, persons of the same earning potential may be taxed more if they are more willing to bear risk, to save, etc. However, this imperfection applies also to the use of measured income for the purpose of preferential treatment and hence does not affect my argument. In general, to the extent that a better index is available for use as a basis of preferential treatment, it can also be used for the tax-subsidy purpose. Unless there are asymmetrical transaction costs (see Part B above), no qualification to my central argument is necessary.

From the discussion above, it may be concluded that none of the complications seem to change my central argument significantly.

V. Concluding Remarks

Changes satisfying my proposed welfare criterion may be called quasi-Pareto social improvements as all relevant (usually income) groups of individuals are made better off in the sense of fulfilling the Kaldor-Hicks-Scitovsky double compensation test within each group. It is also only *quasi* Pareto in the sense that repeated applications of the Kaldor-Hicks-Scitovsky criterion may lead to cyclicity and Pareto inferiority (Appendix, part A). This is due to the necessary approximate nature of all objective measures of subjective welfare changes. But the discrepancies are likely to be overwhelmed by inaccuracies in data collection for changes with effects thinly spread across a large number of individuals. If we are quite certain of the fulfillment of our criterion despite data inaccuracies, Pareto inferiority and inconsistency will almost certainly be absent. One or two possible exceptions may be regarded as an unavoidable cost of using a generally good rule in practice (Appendix, part A).

When combined with the third-best equality-incentive argument, my criterion leads to the forceful conclusion of a dollar is a dollar irrespective of income groups. This conclusion seems to make the criterion itself redundant. The apparent paradox is easily explained. A dollar is a dollar is evaluated prior to disincentive effects while my welfare criterion is applied to the final outcome inclusive of any disincentive effects. But the principle of a dollar is a dollar does provide a powerful simplification in economic assessment of any policy, change, etc. in general, and in cost-benefit analysis in particular. We may assess on pure efficiency considerations unless certain specific preferential policies may be justified on grounds of asymmetrical transaction costs or asymmetrical political feasibility, or the like (Section IV). Pure equality considerations do not provide adequate justification for preferential treatment between the rich and the poor, or for a departure from a dollar is a dollar.

Consider again the water pricing example of Section II. Unless some specific argument (on top of equality) can be advanced for cross subsidization, it is better to go further than the proposal of Section II which maintains the present cross subsidization. It is better to have a straight dollar-for-dollar pricing system without cross subsidization. One possible argument for cross subsidization is the infeasibility of making the income tax system more progressive to compensate for the withdrawal of cross subsidization. This is quite likely true in the short run, but not in the long run (Section IV). Another possible argument is that it may be more difficult to avoid paying the cross subsidies attached to properties than to avoid paying income taxes. If this is true, it is still better to have a property tax than attaching the tax to *unused* water entitlements with the efficiency loss of discouraging water conservation. If this is politically infeasible for some reason, the proposal of Section II based on my welfare criterion may then be considered.

edge of individual input supply functions that they recognize to be not available. But they suggest that some rough discrimination between broad groups of income earners such as doctors vs. ditch diggers may yet be feasible. But if such a discrimination of supplementing "wage rates for the one and limit[ing] wages for the other" (p. 522) is feasible, there is little reason to expect that it is not feasible to have higher income tax rates for doctors and lower for ditch diggers.

APPENDIX

A. *The Acceptability of the Kaldor-Hicks-Scitovsky Criterion in the Absence of Distributional Considerations*

Here I provide an argument in favor of accepting the Kaldor-Hicks-Scitovsky criterion, *in the absence of distributional consideration*, as a *practical* criterion for general economic application, despite the acknowledged possibility of cyclicity after repeated usage.

Even abstracting away distributional considerations (or other grounds for differentiating "deservingness"), the Kaldor-Hicks-Scitovsky criterion (which requires the satisfaction of both the Kaldor and Hicks compensation tests) is not an ideal sufficient condition for a social improvement. This is so since a logical contradiction is possible after *repeated* applications of the criterion (W. M. Gorman, 1955; John Chipman and James Moore, Section 3; my book, p. 66). As illustrated in the utility space of Figure 3 (where the curves are utility possibility curves), the changes from q^1 to q^2, q^2 to q^3, q^3 to q^4, and q^4 back to q^1 all satisfy the criterion. Moreover, q^4 is in fact Pareto inferior to q^1. Thus the criterion is not only logically inconsistent, but can lead to a Pareto-inferior situation.

The possible cyclicity and Pareto inconsistency of the Kaldor-Hicks-Scitovsky criterion illustrates the difficulty of making social welfare judgments without interpersonal comparison of individual cardinal utilities. In fact, I have established elsewhere (1982, Proposition 4) that, given a sufficiently wide domain (a condition less demanding than Universal Domain and satisfying conventional "economic" assumptions of self interest, nonsatiation, etc.), there exists no non-cardinalistic ranking rule yielding consistent ordering of social states satisfying anonymity and the Pareto principle. A non-cardinalistic ranking rule is a rule specifying social ranking of pairs of social states based on individual *rankings* (but not on preference intensities) of the respective pairs and (possibly) on some (but not all) objective characteristics of the social states. The

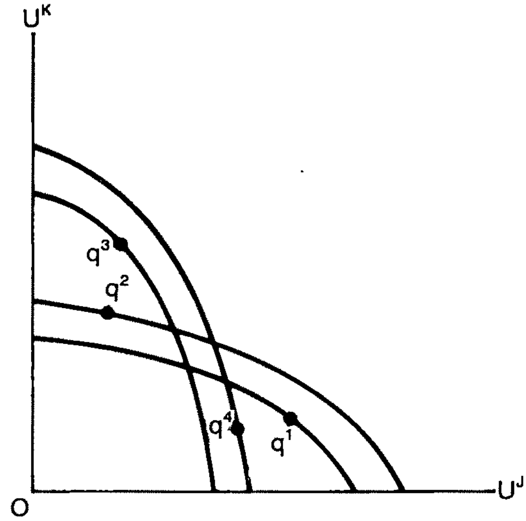


FIGURE 3

Kaldor-Hicks-Scitovsky criterion is in fact more than a non-cardinalistic ranking rule as it effectively uses the amount of compensation required and the willingness to pay as indirect measures of subjective preference intensities. But since these indirect measures are not perfect in their correspondence with subjective preferences, inconsistencies may arise. This difficulty is present for all nonsubjective measures of welfare. Since the relation between units of any external yardstick of welfare such as money and internal (subjective) units of welfare is in general not a constant (making intersections of utility possibility curves as in Figure 3 possible), such objective measures can be, by their very nature, no more than an approximate measure of welfare, even abstracting from the problem of inaccuracies in practical data collection. If we recognize the necessity for this approximate nature, the possibility of inconsistency (as well as such problems as path-dependency in consumers' surplus measurement) becomes acceptable unless the discrepancies involved are substantial and frequent. For general application in judging the desirability of economic policies with widely and thinly spread costs and benefits, the discrepancies involved are likely to be small and largely offsetting to each other. Thus, if the satisfaction of the Kaldor-Hicks-

Scitovsky criterion is fairly certain despite inaccuracies in data collection, we can be quite sure that the problem created by the approximate nature of our measurement will be overwhelmed. One or two odd cases of inconsistencies may still remain, but that can be regarded as an unavoidable cost of using a generally desirable rule. For example, the rule that motorists must stop at red traffic lights is certainly desirable as it prevents accidents and congestion, but it also creates some unnecessary waiting time. It is Pareto inferior for someone to wait for a green light when no one is crossing the intersection from any direction. But it is impractical to allow such Pareto improvements of crossing red lights without creating dangerous accidents. Similarly, before it is practicable to use some direct measurement of subjective welfares,[9] we have to make do with imperfect substitutes usually in some form of willingness to pay. Recognizing the necessary approximate nature of such measures, the possibility of inconsistency is not a compelling objection. The use of the Kaldor-Hicks-Scitovsky criterion, when distributional considerations have been dealt with as suggested in Section I, can thus be justified.

B. *Dealing with Changes in Income-Group Status*

In Section I, a welfare criterion (or a sufficient condition for a social improvement) is proposed that requires the satisfaction of the Kaldor-Hicks-Scitovsky (double) compensation test for each income group. A problem arises when some persons in an income group are made so much better off or worse off as to change their status into different income groups. For example, if a change affects the income levels of the poor by making most of the poor poorer and making a few rich such that compensation is possible, one may not wish to regard such a change as socially desirable. My compensation test will work best if no such changes of income status take place. But since a person on the top (bottom) of an income group need only to be made a little richer (poorer) to change his status, a change in status by not more than one step may be regarded as acceptable such that he can be included in his original income group for the purpose of compensation test. If the social marginal utility of a dollar is regarded as approximately equal for all individuals in the same income group, they do not differ by too much for neighboring income groups.

For persons who jump income groups by more than one step, the following method of using my compensation test is suggested. One who jumps up the scale from income group G to $G+A$, where A is a positive integer larger than one, may be included in his original income group (G) for testing compensation provided his gain is (hypothetically) reduced to an amount making him no higher than the top of the income group $G+1$. In other words, his gain in income above this amount is disregarded for the purpose of the compensation test. To be conservative, we may insist that no one be made to jump down the income groups by more than one step. Alternatively, we may require that if such a downward jumping occurs, the person involved should be included in his new income group for the purpose of the compensation test. If we adopt the latter alternative, it is consistent to allow the following. For a person who jumps upward by more than one step, if his original income group already satisfies the compensation test without including him, his gain may be included, if needed, in the compensation test of the income group he moves into.

C. *A Dollar is a Dollar: Proof of Proposition* 1

To establish Proposition 1 in Section III, let us adopt the following simplifying assumptions: 1) there is no political constraint on redistribution through taxation; 2) the administrative costs of the pure taxation system are no higher than its alternative, for any same degree of equality attained; 3) all individuals know the relevant taxation scale, details of government expenditure, etc.; 4)

[9] Such as the psychological measurement based on just noticeable differences; see my article (1975) on such a measurement and ways to overcome practical difficulties.

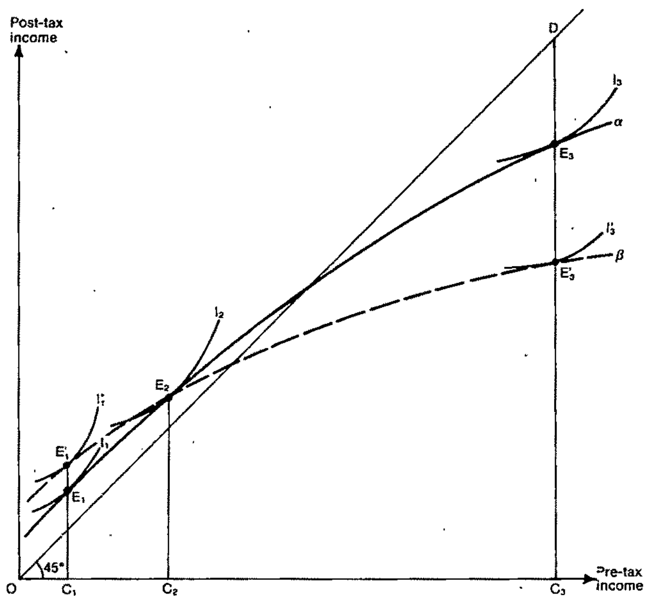


FIGURE 4

there are no money illusion or similar "irrational" preferences. In Section IV, we see that the relaxation of these assumptions does not affect the argument significantly.

Let us define a system of perfect preferential treatment as one that involves a degree of preferential treatment that is monotonically decreasing in incomes. Initially, I shall establish the proposition by assuming, first, that system a is perfect, and second, that individuals have the same utility function between income and leisure (but may have different earning abilities).

Consider Figure 4 where curve α represents a given income tax schedule relating post-tax to pre-tax income levels. For example, a person earning $0C_3(=C_3D)$ will be taxed DE_3 and left with E_3C_3 as his post-tax income. Each person has a given income-earning ability. Subject to this earning ability, each person may choose different levels of pre-tax income by varying his hours and intensity of work. His choice depends, of course, on his subjective preference (with respect to leisure, consumption, and the preferential system), the tax schedule, and the system of preferential treatment. Let alternative A be the tax schedule α and a given system of preferential treatment a. Even with the assumption that all individuals have the same subjective preference (or utility function), persons of different earning abilities may have different indifference maps as defined on Figure 4. (This is similar to James Mirrlees' model of optimal income taxation. See J. K. Seade, 1977, for a diagrammatical illustration.) Given some mild assumptions, income varies positively and continuously with earning ability (Mirrlees, 1971). Geometrically, a person with higher earning ability has a flatter indifference curve at a given point. This is so since a person of lower earning ability needs to work more to earn a given income. The equilibrium points (E_1, E_2, E_3) of three individuals under alternative A are depicted in Figure 4.

Now let us dismantle the system of preferential treatment a. This will make the rich better off and *may* make the poor worse off. If system a is so inefficient such that its dismantling makes everyone better off, we have a stronger case for its removal. Thus, let us take the case where the poor will be made worse off by its removal. Since the preferential treatment is assumed perfect, there exists an intermediate income level (say C_2) at which the individual would stay indifferent by the removal of system a. This individual must exist in a model of a continuous distribution of individuals but may not exist in the discrete case. But the actual existence of this individual is of no consequence to my

argument. After the removal of system a, the new indifference curve of this individual that corresponds to the same level of utility as I_2 must still pass through the point E_2. On the other hand, individual 3 who is made better off by the removal of system a, must have a new indifference curve (I_3'), that corresponds to the same level of utility as I_3, passing through a lower point E_3'. With no preferential treatment against him, he now only needs a lower level of post-tax income to attain the original utility level at the same level (OC_3) of pre-tax income.[10] Conversely, the new indifference curve (I_1') for individual 1 passes through a higher point E_1'. Tracing through all points such as E_1', E_2, E_3', we arrive at the new tax schedule β. With the usual continuity assumptions, this schedule will also be continuous and smooth. (Continuity and smoothness are not really necessary for my central proposition but they make the illustration easier and enable it to be put in terms of the familiar tangency condition for maximization.) Let us call the tax schedule β with no preferential treatment alternative B. If the government can collect at least as much revenue as before to maintain public expenditure, it is clear that everyone will remain at least as well off as under alternative A. This is so since each person can always choose to earn the same amount of pre-tax income and attain the same level of utility as before. However, if alternative B has greater disincentive effects, many individuals may choose to earn less and government revenue may be smaller than before. Let us examine this possibility.

Consider Figure 5 which is an enlargement of the relevant section of Figure 4. If the new indifference curve of individual 2 does not only pass through E_2 but also stays un-

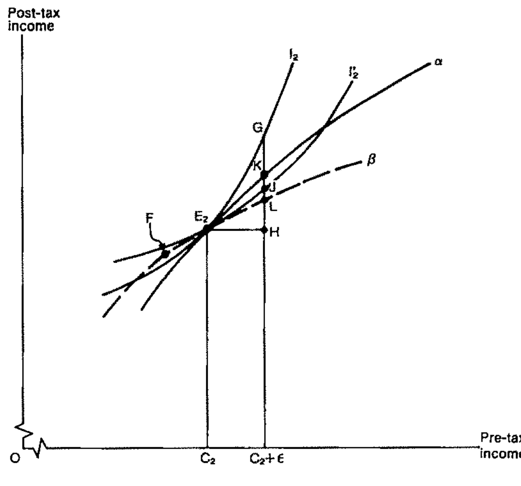


FIGURE 5

changed at I_2 (at least in the neighborhood of E_2), the new tax schedule β, being flatter than α, must cut this indifference curve (I_2). Individual 2 would then choose to earn a lower level of pre-tax income, say, at point F. However, the actual new indifference curve (I_2') must be flatter than the old one I_2 at point E_2. With no preferential treatment, he would need less (more) post-tax income if he were to earn a higher (lower) level of pre-tax income. (Otherwise β would not be flatter than α to begin with.) Moreover, it can be shown that I_2' must be tangential to β at E_2. The tax schedule α touches the (highest) indifference curve I_2. Both I_2 and α are reduced in slope to become I_2' and β, respectively. Moreover, the reduction in slope must be the same at the point E_2. Hence β touches I_2' at E_2. To see this more clearly, consider a slightly higher level of pre-tax income $C_2 + \varepsilon$. The slope of I_2 at E_2 (denoted S_2) may be approximated by GH/E_2H. This approximation will become exact equality as we make ε approach zero, given smoothness in the curve. Similarly, the slope of I_2' at E_2 (denoted S_2') $\simeq JH/E_2H$, and the slope of α at E_2, $S_\alpha \simeq KH/E_2H$, and the slope of β at E_2, $S_\beta \simeq LH/E_2H$. Since the denominators of all these slopes are the same, we may concentrate on the numerators. Slope S_2 is larger than S_2' by (approximately) GJ, ignoring the denominator. GJ measures the extent to which individual 2 would be made better off

[10] This is obviously true with the assumption of given earning abilities. With the relaxation of this assumption, it may be thought that I_3' may pass above E_3 (and I_1' below E_1) if the wage rate of individual 3 (individual 1) is sufficiently increased (reduced) by the operation of preferential system a. If this is so, it means that preferential system a in fact favors the rich rather than the poor when its full effects (including the indirect effect on earning abilities) are taken into account. This possibility may thus be disregarded. It is also clear that the indirect effect is unlikely to outweigh the direct effect.

by the removal of preferential system a had he chosen to earn $C_2 + \varepsilon$ instead of C_2.[11] Slope S_α is larger than S_β by KL. Thus KL measures the extent to which the individual who actually earned $C_2 + \varepsilon$ under alternative A would be made better off by the removal of preferential system a if he were to keep earning $C_2 + \varepsilon$. It must be recognized that GJ need not equal KL. Although the pre-tax $(C_2 + \varepsilon)$ incomes of both individuals are the same if the above hypothetical conditions prevail, they have different earning abilities. The same pre-tax income must then imply different hours or intensities of work. This difference in hours of work may then make them willing to forgo a different sum of post-tax income to remove the same system (and the same degree) of preferential treatment. However, as we make ε approach zero, not only do the above approximate measures of slopes become exact, the difference in the earning abilities of the two individuals also approaches zero. Given continuity, the amount by which S_2 is larger than S_2' will than be equal to the amount by which S_α is larger than S_β at point E_2 as the measures of these slopes are made exact. Since $S_2 = S_\alpha$ at E_2, so $S_2' = S_\beta$ at E_2. It is then not difficult to see that, under alternative B, not only can individual 2 attain the same level of utility as before by earning the same income as before, he has no incentive to earn a different income. Given some convexity assumptions, he would be positively worse off if he operated at a different point.

Let us now go back to Figure 4 to consider the position of individual 3. Now S_3' (slope of I_3' at E_3') may differ from S_3 for two reasons. One is the same as that which makes S_2' differ from S_2, that is, the removal of preferential treatment tends to make the slope flatter. But since E_3 and E_3' are now two different points, there is an additional reason for S_3' differing from S_3. The difference in post-tax income may make the individual have a different tradeoff between consumption (or post-tax income) and leisure (related to pre-tax income). However, S_β (at E_3') also differs from S_α (at E_3) for these two reasons. It is thus not difficult to see that S_β must equal S_3' at E_3'. Individual 3 will choose to earn the same amount of pre-tax income as before. Similar reasoning shows that under alternative B, all individuals will choose to earn the same amounts of pre-tax income as under alternative A. Alternative B thus provides the same degree of incentives and the same degree of equality in the distribution of real income (utility) as alternative A.

Even if all individuals earn the same amount of pre-tax income, can we be sure that government revenue is no smaller under alternative B? In Figure 4, let C_1 (it could be zero) be the lowest pre-tax income earned and C_3 be the highest. The change in government revenue in moving from alternative A to B equals the area $E_2E_3E_3'$ weighted by population density function along the horizontal axis minus the area $E_1'E_1E_2$, similarly weighted. It is clear that the weighted area $E_2E_3E_3'$ must be larger than the weighted area $E_1'E_1E_2$. The former measures the aggregate amount by which all individuals earning more than C_2 are made worse off by preferential system a. The latter area measures the aggregate amount by which all individuals earning less than C_2 are made better off by system a. If the former area was smaller than the latter, system a would be justified on pure efficiency grounds to start with. Thus if system a is truly preferential, the former area must be larger than the latter. For example, the use of unequal income weighting in cost-benefit analysis may sanction projects with positive unweighted aggregate net benefits. But such projects will be sanctioned without the use of unequal income weighting and would be undertaken under alternative B, too. Hence the difference in tax schedules α and β is caused by the effects of preferential measures such as unequal income weighting when they are effective in sanctioning projects with *negative* unweighted aggregate net benefits.

[11] This is so due to the argument in Section III abstracting away the possible second-best effect of preferential system a. This effect may change the tradeoff between consumption and leisure, i.e., the slopes of indifference curves in the figure. Then GJ may partly reflect this second-best effect. However, this second-best effect may go in either direction and, as argued above, would result in a negative expected gain. Thus, by abstracting away the second-best effect, we in fact give alternative A an advantage.

From the above discussion, it can be seen that alternative B not only provides the same degree of incentives and the same degree of equality in the distribution of real income (utility), it also generates more government revenue than alternative A. This extra amount of revenue is a measure of the superiority of B (no preferential treatment but more progressive income taxation) and can be used to make everyone better off by increasing public expenditure and/or lowering taxes all round.

The argument above is based on the assumption of *perfect* preferential treatment. The degree of preferential treatment is taken to be a monotonically decreasing function of incomes across all individuals. In practice, preferential treatment in government expenditure cannot be perfect. For one thing, some expenditures benefit all people in the same geographical area. The government may choose to spend more on poor areas, but a rich person living in a predominantly poor area will benefit as well. A person is not likely to change his place of residence each time his income is increased. Hence, for any person living in a particular area, he will not be appreciably adversely affected if he earns more by those government expenditures that are geographically specific. The increase in his income does not appreciably increase the average income of the whole area. But for the purpose of income taxation, it is individual income alone that counts and not the average income of the whole area. It follows that the disincentive effect of pure income taxation (alternative B) is greater than that of income taxation with lower progressivity but with *imperfect* preferential treatment (alternative A'). Does it follow that alternative B is inferior to alternative A'? No, as the following paragraph shows.

The reason we have to make do with imperfect preferential treatment is the infeasibility or very high costs of effecting perfect preferential treatment, not that we prefer imperfect preferential treatment (alternative A') to perfect preferential treatment (alternative A) as such. Abstracting from the problems of feasibility and transaction costs, alternative A is preferable to A'. But it has been argued above that alternative B is preferable to alternative A (without counting the transaction costs involved in A). It follows that alternative B must be preferable to A'. This is so despite the fact that the disincentive effect is higher under B. The imperfection of alternative A' involves welfare loss in terms of inequity which must be larger than the costs of higher disincentive effects of B or A (which have the same incentives), otherwise A' would be preferable to A. Thus the problem of imperfection in preferential treatment does not affect my conclusion.

REFERENCES

Atkinson, Anthony B. and Stiglitz, Joseph E., "The Design of Tax Structure: Direct versus Indirect Taxation," *Journal of Public Economics*, July–August 1976, *6*, 55–75.

Baumol, William J. and Fischer, Dietrich, "The Output Distribution Frontier: Alternatives to Income Taxes and Transfers for Strong Equality Goals," *American Economic Review*, September 1979, *69*, 514–25.

Bergstrom, Theodore C. and Cornes, Richard C., "Independence of Allocative Efficiency from Distribution in the Theory of Public Goods," *Econometrica*, November 1983, *51*, 753–66.

Boadway, Robin W., "The Welfare Foundation of Cost-Benefit Analysis," *Economic Journal*, December 1974, *84*, 926–39.

Bös, Dieter, "Income Taxation, Public Sector Pricing, and Redistribution," typescript, 1984.

Buchanan, James M. and Tullock, Gordon, *The Calculus of Consent*, Ann Arbor: University of Michigan Press, 1962.

Chipman, John S. and Moore, James C., "The New Welfare Economics 1939–1974," *International Economic Review*, October 1978, *19*, 547–84.

Corden, W. M., *Trade Policy and Economic Welfare*, London: Oxford University Press, 1974.

Dodgson, J. S., "Consumer Surplus and Compensation Tests," *Public Finance*, No. 3, 1977, *32*, 312–20.

Feldstein, Martin S., "Distributional Preferences in Public Expenditure Analysis," in H. M. Hochman and G. E. Patterson, eds., *Redistribution Through Public Choice*, New

York: Columbia University Press, 1974.

Gorman, W. M., "The Intransitivity of Certain Criteria used in Welfare Economics," *Oxford Economic Papers*, February 1955, *7*, 25–35.

Harberger, A. C., "The Three Basic Postulates for Applied Welfare Economics: An Interpretive Essay," *Journal of Economic Literature*, September 1971, *9*, 785–97.

______, "On the Use of Distributional Weights in Social Cost-Benefit Analysis," *Journal of Political Economy*, April 1978, *86*, S87–S120.

Hicks, John R., "The Valuation of Social Income," *Economica*, May 1940, *7*, 105–24.

______, "The Rehabilitation of Consumer's Surplus," *Review of Economic Studies*, February 1941, *8*, 108–16.

Hirshleifer, Jack, de Haven, J. C. and Millman, J. W., *Water Supply: Economic Technology, and Policy*, Chicago: University of Chicago Press, 1960, rev. 1969.

Hotelling, Harold, "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," *Econometrica*, July 1938, *6*, 242–69.

Kaldor, Nicholas, "Welfare Propositions of Economics and Interpersonal Comparison of Utility," *Economic Journal*, September 1939, *49*, 549–52.

Kemp, Murray C. and Ng, Yew-Kwang, "On the Existence of Social Welfare Functions, Social Orderings and Social Decision Functions," *Economica*, February 1976, *43*, 59–66.

______ **and** ______, "More on Social Welfare Functions: The Incompatibility of Individualism and Ordinalism," *Economica*, February 1977, *44*, 89–90.

______ **and** ______, "The Incompatibility of Individualism and Ordinalism," *Mathematical Social Sciences*, July 1982, *3*, 33–38.

Leibenstein, Harvey, "Long-Run Welfare Criteria," in J. Margolis, ed., *The Public Economy of Urban Communities*, Baltimore: Johns Hopkins University Press, 1965.

Little, I. M. D., *A Critique of Welfare Economics*, London: Oxford University Press, 1950; 1957.

Mirrlees, James A., "An Exploration in the Theory of Optimal Income Taxation," *Review of Economic Studies* April 1971, *38*, 175–208.

______, "Optimal Tax Theory: A Synthesis," *Journal of Public Economics*, November 1976, *6*, 327–58.

______, "The Theory of Optimal Taxation," in K. J. Arrow and M. D. Intriligator, eds., *Handbook of Mathematical Economics*, Vol. III, Amsterdam: North-Holland, 1981.

Mishan, Ezra J., *Welfare Economics: An Assessment*, Amsterdam: North-Holland, 1969.

Musgrave, Richard A., "Cost-Benefit Analysis and the Theory of Public Finance," *Journal of Economic Literature*, September 1969, *7*, 797–806.

Ng, Yew-Kwang, "Bentham or Bergson? Finite Sensibility, Utility Functions, and Social Welfare Functions," *Review of Economic Studies*, October 1975, *42*, 545–70.

______, "Towards a Theory of Third Best," *Public Finance*, No. 1, 1977, *32*, 1–15.

______, *Welfare Economics: Introduction and Development of Basic Concepts*, London: Macmillan, 1979; 1983.

______, "Beyond Pareto Optimality: The Necessity of Interpersonal Cardinal Utilities in Distributional Judgements and Social Choice," *Zeitschrift fur Nationalökonomie*, November 1982, *42*, 207–33.

Polinsky, A. Mitchell, "Probabilistic Compensation Criteria," *Quarterly Journal of Economics*, August 1972, *86*, 407–25.

Sandmo, Agnar, "Optimal Taxation: An Introduction to the Literature," *Journal of Public Economics*, June 1976, *6*, 37–54.

Scitovsky, Tibor, "A Note on Welfare Propositions in Economics," *Review of Economic Studies*, November 1941, *6*, 77–88.

Seade, J. K., "On the Shape of Optimal Tax Schedules," *Journal of Public Economics*, April 1977, *7*, 203–35.

Skolnik, M. L., "A Comment on Professor Musgrave's Separation of Distribution from Allocation," *Journal of Economic Literature*, June 1970, *8*, 440–42.

Suzumura, Kotaro, "On Distributional Value Judgments and Piecemeal Welfare Criteria," *Economica*, May 1980, *47*, 125–39.

Tobin's *q* and the Structure-Performance Relationship

By Michael Smirlock, Thomas Gilligan, and William Marshall*

The "traditional" interpretation of the structure-conduct-performance (*S-C-P*) relationship is based on the proposition that market concentration fosters collusion by firms. One result of this collusion is that monopoly rents are achieved by firms in the industry. In contrast, the "efficient structure" interpretation asserts that concentration emerges from competition and is the result in industries characterized by firms possessing an advantage in production. As a result of their efficiency, leading firms earn economic or Ricardian rents.[1]

The purpose of this paper is to reexamine the empirical relationship between firm rents and market structure in order to assess the validity of these hypotheses. Few studies test these hypotheses simultaneously (for example, Harold Demsetz, 1973; John Carter, 1978) and these papers either rely on accounting measures of rents or do not adequately control for differences in firms' market shares. By explicitly introducing market share and concentration into our tests we are able to better assess the validity of the competing hypotheses. Another feature of this study is the use of Tobin's q (hereafter q) to measure firm rents. Defined as the ratio of market value to replacement cost of the firm, q has been shown by Eric Lindenberg and Stephen Ross (1981) to provide an accurate measure of the capitalized value of the rents attributable to both monopoly and firm-specific factors. This measure avoids many of the deficiencies attributable to accounting profit rates as measures of firm rents (for example, Yale Brozen, 1970; Stanley Ornstein, 1972; George Stigler, 1963; Franklin Fisher and John McGowan, 1983).

In this paper, Section I discusses the theories, summarizes the relevant literature, and describes the methodology employed to discriminate between the two hypotheses. The key issue is the extent to which firm rents are due to superior efficiency vis-à-vis market concentration. The value of any *S-C-P* test, therefore, partially hinges on the adequacy of the proxy for rents. In Section II the advantages of q as a measure of firm rents relative to accounting profit rates are described. In Section III the data are described and the empirical tests that we undertake are presented. The results of these tests provide support for the efficient structure hypothesis over the traditional hypothesis. Section IV contains a summary of the paper and our conclusions.

I. Firm Rents and Market Structure

The vast majority of empirical *S-C-P* studies have reported a weak positive relationship between market concentration and firm profit rates. The traditional hypothesis (for example, Joe Bain, 1951; William Shepherd, 1972; Leonard Weiss, 1974) interprets this relationship as evidence that concentration induces supracompetitive pricing behavior by lowering the cost of collusion. Consequently, firms in concentrated industries earn monopoly rents. Further, concentration is not associated with the superior efficiency of the leading firms, so that the benefits of collusion accrue equally to all firms in the industry. Combined with the proposition that the leading firms in concentrated industries tend to be larger than the minimum efficient scale, this interpretation provides a basis for the structuralist antitrust policy of deconcentration. The result of such a policy would be to lower prices with no sacrifice in efficiency.

*Assistant Professor of Finance, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104; Assistant Professor of Economics, California Institute of Technology, Pasadena, CA 91125; and Associate Professor of Finance, School of Business Administration, Washington University, St. Louis, MO 63130, respectively. We thank Peter Linneman, David Qualls, Jess Yawitz, Harold Demsetz, and the referees for their comments.

[1] For excellent presentations of the competing hypotheses see Harvey Goldschmid, M. Michael Mann, and J. Fred Weston (1974), and Yale Brozen (1982).

Conversely, the efficient structure hypothesis (Demsetz, 1973; Brozen, 1970, 1982; John McGee, 1974; Sam Peltzman, 1977) asserts that market concentration is not a random event but rather the result in industries where some firms possess superior efficiency. Firms with a comparative advantage in production become large and obtain a high market share and, as a consequence, the industry becomes more concentrated. These firms earn economic or Ricardian rents (and, if collusion also results, monopoly rents). Demsetz (1973, 1974) has noted that most *S-C-P* studies focus on large firms with high market shares. He asserts that the correlation found between concentration and profits is spurious and attributable to the interrelationship between market share, concentration, and superior efficiency. One implication of this hypothesis is that a policy of deconcentration may reduce prices, but any benefits may be more than offset by the adverse effect on costs. In the case where collusion is not facilitated by concentration, an antitrust policy that promotes deconcentration would result in higher costs with no compensation in the form of lower prices.

There are few studies which simultaneously test these hypotheses and those that do contain methodological shortcomings that make their conclusions tenuous. Demsetz (1973) based his test on the supposition that if both large firms and small firms in concentrated industries achieve rents, then collusion is present and the traditional hypothesis cannot be rejected. If only leading firms achieve rents, however, then the efficient structure hypothesis is supported, since small firms receive no benefits from concentration. In implementing this test, Demsetz utilized three-digit IRS data to calculate concentration-profitability correlations across industries for each of five asset size classes. Finding a positive correlation only for the largest firms, Demsetz concludes that large firms are more efficient than their smaller rivals. Carter, noting the weaknesses of IRS data and the asset size delineations chosen by Demsetz, utilized regression analysis to examine whether leading firms in an industry have a higher price-cost margin than secondary firms. He finds this to be the case and interprets this evidence as corroborating Demsetz's research. Weiss, however, has argued that the Demsetz-Carter findings are consistent with various theories of oligopoly, which suggest that the leading firms reap the main advantages of concentration. Given this, proponents of the traditional hypothesis view these empirical studies as providing no evidence either to substantiate the efficient structure hypothesis or to refute the traditional hypothesis.

Weiss suggests that the correct test of the competing hypotheses would be one that "takes both market share and concentration into account at the same time" (pp. 225–26). A straightforward way to do this is to estimate an equation with an appropriate measure of firm rents as the dependent variable and both market share and concentration included as independent variables. Weiss notes that several studies have done this and cites two (Bradley Gale, 1972; Blake Imel and Peter Helmberger, 1971) in particular as supporting the traditional hypothesis.

Closer inspection of these studies, however, suggests that any such conclusions are premature. Gale claims that profit rates are higher, given any market share, "when the competitive environment is characterized by high concentration..." (p. 422). The support for this statement appears to be a positive and significant coefficient on an interaction variable between market share and a high concentration dummy variable.[2] At least three criticisms of Gale's analysis and conclusions are relevant for our work. First, his equation, which includes concentration variables and interaction terms between market share and concentration, does not significantly increase explanatory power over an equation that simply employs a continuous measure of market share as its sole market structure variable. Second, the coefficient on the high concentration dummy variable in the more complex formulation is not significant. Only the *interaction* between market share and high concentration is significant. Since firms with high market share are prob-

[2] See Gale (Tables 1 and 2, pp. 416 and 418, respectively).

ably in highly concentrated markets, the interaction variable may be proxying for the omitted market share variable. Third, the difference between the interaction of market share with low concentration and market share with high concentration is insignificant, further indication that this approach simply proxies for the omitted market share variable.

The findings of Imel and Helmberger are also suspect for methodological reasons. They define the profit rate as total after-tax profits less an *ad hoc* adjustment for implicit equity return divided by total sales. This measure is without any apparent theoretical justification.[3] Further, the market share variable used is actually firm sales relative to the total sales of the four largest producers, not total market sales. The dubious measures of profit rate and market share may be responsible for their results.

Stavros Thomadakis (1977) also examines the *S-C-P* relationship utilizing both market share and concentration in the same regression equation. He does not, however, consider the two different hypotheses but rather views his tests as an extension of the traditional hypothesis. Although he finds that both market structure variables are significant, his measure of market rents is seriously flawed. Specifically, while asserting (incorrectly) that monopoly rents can be measured as deviations of market value from replacement cost, Thomadakis utilizes book value rather than market value of debt, and book value rather than replacement value of assets in calculating his measure of rents. In addition, he arbitrarily scales the market-asset value difference by sales.[4]

More recently, Weiss and George Pascoe (1981) and David Ravenscraft (1983) have utilized data derived from the Federal Trade Commission's Line of Business survey. These authors found no significant relationship between profits and concentration when market share and concentration are included in the same regression equation. Based on this, Ravenscraft concludes that the observed profit-concentration relationship in his industry regression analysis reflects advantages that larger firms enjoy over smaller firms. The dependent variable in these studies, however, is defined as operating income divided by sales, which is at best a crude measure of a price-cost margin. Further, the empirical definition of income requires accurate allocation of costs, which is especially difficult and suspect for a single product of a multiproduct firm such as those in the FTC survey. Thus, the measure of rents used in these papers is subject not only to the criticisms of accounting profits in general, but suffer from additional shortcomings due to the disaggregate nature of the data.

The methodology used in this paper is similar to that used by Ravenscraft (1983) and is based on Weiss's assertion regarding the appropriate test of the alternative hypotheses. Specifically, we estimate the equation

$$(1) \quad R = a_0 + a_1 MS + a_2 CR + \sum_{i=3}^{n} a_i Z_i,$$

where R is a measure of firm rents, MS is firm market share, CR is market concentration, and Z is a vector of control variables frequently employed in the literature. The usefulness of (1) in discriminating between these hypotheses is straightforward. A coefficient of $a_1 = 0$ implies that, *ceteris paribus*, MS does not affect R and suggests that all accrued rents are monopoly rents due to market concentration. The finding $a_1 = 0$ is thus consistent with the traditional hypothesis. On the other hand, $a_1 > 0$ would imply that firms with high market share are more efficient and earn rents because of this efficiency. The finding $a_1 > 0$ is thus consistent with the efficient structure hypothesis. Similarly, the traditional hypothesis implies that $a_2 > 0$ since concentration will result in monopoly rents that will not be attributable to MS. If the efficient structure hypothesis is

[3] This criticism is in addition to those discussed below in Section III, which also apply to the profit/equity measure used by Gale.

[4] It should be noted that Thomadakis' rent measure is similar in intent to q. In computing his rent measure, however, he calculates neither market value nor replacement cost of capital of the firm. Further, his analysis ignores the possibility that rents can accrue for reasons other than monopoly power.

the more accurate description of market behavior, then $a_2 = 0$ since market concentration will not impact on firm rents.[5]

Thus, $a_1 > 0$, $a_2 = 0$ provides support for the efficient structure hypothesis while $a_1 = 0$, $a_2 > 0$ provides support for the traditional hypothesis. This, of course, does not exhaust the possible combinations of a_1 and a_2. Also of interest is the case where both a_1 and a_2 are greater than zero. Advocates of the traditional hypothesis would interpret such a result as demonstrating that all firms in concentrated markets earn monopoly rents from collusion and that these benefits, as suggested by theories of oligopolistic behavior, are distributed unevenly with the larger firms in the market capturing the lion's share of monopoly rents. Proponents of the efficient structure hypothesis, on the other hand, would view such a finding as evidence that leading firms are more efficient than their rivals and that market concentration fosters collusion that results in some monopoly rents being earned. Alternatively, these same proponents may argue that a given market share (of a superior firm) has a different meaning depending on the size of its rivals. Concentration may track some of the differences in the degree of superiority and thus may be positively related to firm rents without suggesting any collusion or monopoly rents.

II. Profits, q, and Measurement of Firm Rents

Most studies of the *S-C-P* relationship have utilized accounting profit rates to measure firm rents.[6] The use of profit rates, typically calculated as the ratio of accounting profits to book equity, as measures of monopoly rents has been extensively criticized. Distortions caused by the tax laws, arbitrary accounting conventions, and the possibility that short-run profit rates reflect disequilibrium conditions have all been cited as empirical problems that may vitiate the results of previous studies. In addition, Stigler has argued that the application of arbitrary discount rates in an attempt to normalize accounting profit rates for firm risk introduces biases of a qualitatively and quantitatively unknown nature. Fisher and McGowan have concluded that the list of restrictions necessary for the coincidence of accounting profits and economic rents is so lengthy that the use "of accounting rates of return to make inferences about monopoly profits is a baseless procedure" (p. 89).

Future firm rents, on the other hand, will be appropriately capitalized by an efficient capital market.[7] By combining financial market data with accounting data, a more accurate measure of firm rents can be derived.[8] Tobin's q provides the framework for this construct. Relying on capital markets to value rents avoids or substantially mitigates most of the shortcomings inherent in accounting profit rates. Among its advantages, capital market valuation appropriately incorporates firm risk, corresponds to an equilibrium valuation of rents and minimizes any distortions introduced by tax laws and accounting conventions. The total market value of the firm consists of three components: the capitalized value of rents due to monopoly power; the capitalized value of rents attributable to scarce factors of production; and the present value of the firm's existing capital stock. q is defined as the ratio of the current market value of the firm to the market value of its productive assets. As such, q provides a precise bound on the monopoly and firm-specific rents of the firm. q will exceed unity by the capitalized value of these rents.[9]

[5]Ravenscraft (1981) has demonstrated that if the effect of market share on profitability reflects efficiency, then concentration will yield an unbiased estimate of collusion when market share is also included as an explanatory variable.

[6]For a detailed review of this literature, see Weiss.

[7]Some authors (for example, Timothy Sullivan, 1977) have used the market equity value to book equity value as a measure of capitalized excess profits. This measure is quite crude in that it fails to capture the significant portion of firm value comprised by debt and, as Weiss notes, would be different for two firms with the same amount of capitalized rents but different capital needs.

[8]The use of capital market data to evaluate issues in industrial organization is a relatively recent phenomenon, but one that is valuable and becoming increasingly widespread. For a literature review, see G. William Schwert (1981).

[9]The q forms a precise bound on rents for firms characterized by positive investments. For firms in a dying industry, or where technical progress has made

TABLE 1—SUMMARY STATISTICS

Variable	Mean	SD	Fractiles of the Sample Distribution of the Statistic[a] Min	.10	.25	.50	.75	.90	Max	Shepherd[b] Mean	SD
q^*	2.30	1.62	.82	1.14	1.30	1.69	2.79	4.43	11.17	–	–
MS	24.35	14.09	5	10	15	20	30	40	80	20.92	12.68
CR	65.92	17.39	25	40	55	65	80	90	98	63.75	17.00
SIZE	6.35	1.18	3.04	4.90	5.57	6.20	7.12	8.03	9.49	–	–
MSG	1.01	0.19	.40	.86	1	1	1	1.19	2.33	–	–

Note: A dash indicates not reported; *MSG* is defined as the ratio of the firm's market share in 1968 to the firm's market share in 1961; *SD* is standard deviation.
[a]This study: $N=132$.
[b]Shepherd: $N=231$.

III. Data, Empirical Tests, and Results

The data utilized in this study are derived from two primary sources. Shepherd provided his data for a sample of 231 manufacturing firms listed in the *Fortune Directory* for the period 1960–69. Data for each firm include market share (*MS*), four-firm concentration ratio (*CR*), market share growth over the time period (*MSG*), and binary variables indicating whether the firm's industry is characterized by low (*LBTE*), moderate (*MBTE*), or high (*HBTE*) barriers to entry. These data are described in detail in Shepherd. These variables are representative of the list of regressors used in *S-C-P* studies, although they are by no means exhaustive.

The Shepherd data set is well suited to our purpose. It consists almost entirely of leading firms and thus, if the efficient structure hypothesis is true, it ought to be evident in this sample. Further, these data have been used in other studies of the concentration-profitability relationship and can be considered representative of most of the data samples that have provided evidence of a relationship between concentration and profits.

The q ratios were calculated for all manufacturing firms included in the Standard and Poor's 400 for 1980 for which 1) there were sufficient data on Standard and Poor's Compustat tape to permit computation for the 1961–69 time period, and 2) complete data from the Shepherd data set were available. These criteria resulted in a total sample of 132 firms. For each firm, the variable q^* is defined as the arithmetic mean of q over the 1961–69 time period.[10] The method used to calculate q follows Lindenberg and Ross and is described in detail in the Appendix.

Summary statistics for the continuous variables used in this study are presented in Table 1 along with the mean and standard deviation, where appropriate, reported by Shepherd for his entire sample. The market structure characteristics of the subset of firms used in this analysis are virtually identical to the entire sample. Also, the mean of q^* is above unity while the range is .82 to 11.17. This suggests that at least some of the firms in the sample are earning rents that have been capitalized into their market value.

As a first step, we regress q^* on each of the market structure variables separately by the equations:

$$(2) \quad q^* = b_{10} + b_{11}MS + b_{12}MBTE + b_{13}HBTE + b_{14}MSG;$$

$$(3) \quad q^* = b_{20} + b_{21}CR + b_{22}MBTE + b_{23}HBTE + b_{24}MSG.$$

If q^* is an appropriate measure of rents then

the firm's capital stock obsolescent, q forms a lower, and not a precise, bound on rents.

[10]We could not match Shepherd's 1960–69 period exactly since the Compustat data tape used in this study contains 20 years of firm-specific data from 1961–80.

TABLE 2—REGRESSION RESULTS

Equation	Intercept	*MS*	*CR*	*MBTE*	*HBTE*	*MSG*	*CR**	*SSE*	*F*
(4.1)	-2.33^{a}	$.061^{a}$		.352	.571	2.79^{a}		208.75	21.03
	(3.35)	(7.45)		(1.27)	(1.60)	(4.77)			
(4.2)	-2.44^{a}		$.037^{a}$	.104	.215	2.16^{a}		258.72	10.84
	(2.96)		(4.49)	(.32)	(.48)	(3.30)			
(4.3)	-2.60^{a}	$.055^{a}$	.009	.261	.380	2.67^{a}		206.91	17.06
	(3.51)	(3.61)	(1.05)	(.90)	(.95)	(4.50)			
(4.4)	-2.31^{a}	$.052^{a}$		.308	.353	2.88^{a}	1.23^{a}	189.47	20.95
	(3.49)	(6.29)		(.88)	(1.33)	(5.14)	(3.57)		

Note: t-statistics are shown in parentheses below coefficients; $N=132$ for all equations.
[a] Coefficient significant at the 5 percent level.

there should be a significant positive relationship between q^* and the market structure variables (MS and CR). The results of this estimation are presented in Table 2. Both b_{11} and b_{21} are significant and positive. Such a relationship is indicative of the similarity of the data set and results of this paper and those reported by Shepherd. Therefore, any difference in findings regarding the disparate hypotheses is not likely to be due to significant differences in the data. If left at this point, our findings and implications would confirm those of most previous studies. Such results, however, shed no light on the validity of the competing hypotheses.

To test the traditional and the efficient structure hypotheses simultaneously, we estimate

$$(4)\quad q^* = d_0 + d_1MS + d_2CR + d_3MBTE + d_4HBTE + d_5MSG.$$

According to the hypotheses described in Section II, estimates of d_1 greater than (equal to) zero and d_2 equal to (greater than) zero are consistent with the efficient structure (traditional) hypothesis but are inconsistent with the traditional (efficient structure) hypothesis. If both d_1 and d_2 are greater than zero, neither hypothesis can be rejected.

The results of estimating (4) are reported in Table 2. Prior to consideration of the main hypotheses, the effect of the control variables deserves some attention. MSG has a positive and significant coefficient. One possible interpretation of this is that MSG is the basis for investors' expectations about which firms are most efficient and will therefore increase their market share over time. Neither $MBTE$ or $HBTE$ is significant, which suggests that barriers to entry, at least as measured by Shepherd, either do not exist, are not a source of extant rents, or that the rents were dissipated in obtaining these entry barriers.

Turning to the market structure variables, note that d_1 is positive and significant at the 1 percent level. The coefficient d_2, however, is not significant at any conventional level. These results support the efficient structure hypothesis, and suggest that once market share is controlled for, concentration does not affect firm rents.[11,12] Firms with large

[11] These results are robust with respect to specification and sample selection. Some studies have employed asset size, which we view as endogenous, as an independent variable. Other studies have employed a measure of advertising intensity. Although generally unrelated to the structure-performance hypothesis and more properly considered an endogenous variable, such a measure may help explain q since it can be viewed as an investment in intangible capital. Additionally, Peltzman contends that a firm that is innovative may find it best to decrease its share of the market so that the relevant market share growth variable is the absolute value of the change in market share. Accordingly, (4) was estimated with these additional variables and modifications. In all cases, the coefficient on market share is positive and significant, and the coefficient on concentration is not significant at the 10 percent level. Nor is there any evidence to suggest our results are driven by influential observations. To assess the effects of influential observations on the parameter estimates, the single row deletion techniques suggested and described by David Belsley, Edward Kuh, and Robert Welsch (B-K-W, 1980) were utilized. Eight observations were judged by the critical values recommended in B-K-W to be leverage points and to affect the

market shares appear to earn rents. Whatever market concentration results does not enable these firms to earn additional (monopoly) rents due to collusion.

Several authors (for example, Bain; Weiss; Franklin Edwards, 1977) have maintained that the relationship between concentration and firm rents is not continuous but discrete. That is, there is some critical level of concentration above which firms are able to, implicitly or explicitly, coordinate their activities. To examine this proposition, (4) is reestimated with the variable CR^* replacing CR, where CR^* is equal to one when concentration is above the threshold level and zero otherwise. Since there is no way of knowing the critical level of concentration a priori, we follow several previous studies and use Quandt's "switching of regimes" technique to determine it.[13] This procedure estimated the critical level of the four-firm concentration ratio to be 90 percent. Consequently, CR^* is set equal to one when this concentration measure equals or exceeds 90 percent, and zero otherwise.[14]

The results are presented as equation (4.4) in Table 2. Prior to examination of these results, however, several caveats concerning the existence of a discrete concentration-monopoly relationship should be noted. As Timothy Hannan notes, given a true dichotomous relationship, one would expect the t-statistics achieved with other threshold levels to rise continuously as the level yielding the highest t-statistic is approached. This is not the case for our data. For example, the t-statistics for the threshold concentration levels of 60, 65, 70, 75, 80, and 85 percent (negative t-statistic indicates negative coefficient) are $-.18$, -1.40, .91, 1.79, 1.94 and .99, respectively, which are neither monotonic nor all positive and none are significant at the 5 percent level.[15] The large increase in significance from concentration of 85 to 90 percent is further reason to be suspicious of these results. While these observations do not constitute a statistical test, the discrete concentration-monopoly relationship receives little support from the data. We present the results of this test for completeness, but caution that any conclusions should at most be tentative.

The coefficients on CR^* and MS are positive and significant. This suggests (assuming a discrete concentration-monopoly relationship is appropriate) that firms in these highly concentrated markets may earn monopoly rents. Even in these cases, differential firm

coefficient estimates on MS or CR. These observations were deleted from the sample and (4) reestimated. In no case were the quantitative or qualitative results changed.

Finally, to assess the possibility of degrading collinearity, we follow the methodology suggested by B-K-W and calculate the condition indexes (CI) and variance-decomposition proportions (VDP) of the data. In both a relative and absolute sense, no evidence of degrading collinearity is found. Both the CI and VDP of (4) are similar to those of equations (2), (3), and (4). Additionally, the highest CI associated with (4) is 17.63, well below the critical level of 30 suggested by B-K-W (p. 156) at which further investigation may be warranted. Moreover, the high variance-decompositions (greater than .5) corresponding to this CI are not associated with MS or CR.

[12]Although Tobin's q is superior to profitability as a measure of firm rents, calculation of q is an intricate procedure that may itself introduce errors. If monopoly power is short run and accounting profits are adequate measures of short-run rents, then additional insight and corroboration may be gained by using profit rates to measure firm rents. Accordingly (4) was reestimated using the average return on invested capital for 1961 and 1969, (π^*). The results of this estimation are (t-statistics are shown in parentheses):

$$\pi^* = \underset{(.36)}{-1.14} + \underset{(5.27)}{.177MS} + \underset{(1.48)}{.047CR}$$

$$+ \underset{(2.07)}{5.38MSG} + \underset{(.69)}{.703MBTE} - \underset{(.10)}{.141HBTE}$$

$$SSE = 2115.29, \quad F = 11.72, \quad N = 132.$$

The coefficient on MS is significant at the 1 percent level, while the coefficient on CR is not significant at the 10 percent level. The results are qualitatively similar to those reported in the text. They provide evidence of differential efficiency, but not collusion, among firms.

[13]See, for example, the articles by Edwards and Timothy Hannan (1979).

[14]Without a market share variable in the regression the critical level of concentration was 80 percent and the coefficient on this variable was positive and significant at the 1 percent level. A total of 18 firms were in industries characterized by 90 percent or more concentration.

[15]For all threshold levels the coefficient on market share is positive and significant.

efficiency is extant, as indicated by the positive coefficient on market share. As Peltzman concludes, policy authorities must weigh the benefits of eliminating monopoly rents against the loss of efficiency from breakup of the largest firms. Finally, we view these findings as further evidence to support the efficient structure hypothesis since differential efficiency is still evidenced in the regressions.[16]

IV. Summary and Conclusions

The primary purpose of this paper has been to test the hypotheses found in the industrial organization literature regarding the structure-performance relationship. A secondary purpose has been to empirically demonstrate the usefulness of, and extend, the introduction given by Lindenberg and Ross of Tobin's q in the industrial organization area.

To consider the second purpose first, we have argued that Tobin's q can provide a more appropriate measure of firm rents than more standard measures such as accounting profit rates. q bounds total rents that accrue from either efficiency or monopoly. By relying on market valuation, we avoid many of the shortcomings of accounting rates of return. Further, the use of q is suggestive of the general value of capital market data to investigate issues in industrial organization.

Employing q as our measure of firm rents, we examined the relationships between rents and market share and concentration. Our data source for the market structure variables, developed by Shepherd, has been extensively used in concentration-profit studies and is typical of the samples that have produced evidence frequently cited in support of the traditional hypothesis. Our results are consistent with those of Demsetz, Peltzman, and Carter, and indicate that the efficient structure hypothesis better describes the relationship between market structure and firm behavior. We find that firms with high market shares earn rents not attributable to concentration. There is little evidence to support the contention that concentration induces collusion and generates monopoly rents in any but possibly the most concentrated of markets.

[16] We tested to see if the significance of $CR90$ was due to a nonlinearity in the market share effect by including the MS squared as an additional variable. The coefficient on this variable was negative and insignificant and left other coefficients and their significance levels unchanged. It is possible the $CR90$ effect is due to some omitted relevant variable, but this issue is beyond the scope of this paper.

Appendix

In computing q, we rely primarily on the methodology of Lindenberg and Ross, who describe their computational procedures in detail. q, as previously defined, is the ratio of the firm's market value (of common stock, preferred stock and debt) to the replacement cost of its assets. It is thus necessary to calculate values for each of these components of the q ratio.

The financial claims against the firm were classified as common equity, preferred stock, long-term debt, current liabilities (including that portion of debt maturing within one year), and other liabilities. Year-end market price was used to calculate the value of equity. Preferred stock was assumed to be a perpetuity valued at the average yield reported by Moody's for the year under consideration. Current and other liabilities were valued at book.

Two procedures were used to value debt. For the first 13 years, the outstanding debt was assumed to have been issued in equal amounts over the preceding 20 years and to have had 20-year maturities at issue. For each portion, the coupon rate was assumed to have been Moody's Composite average yield on industrial bonds for the year of issue. For the most recent 7 years, the debt maturing in 2, 3, 4, and 5 years is reported by Compustat. These amounts were assumed to have been issued with 20-year maturities and coupons assigned as above. The remainder of the debt was assumed to have been issued in equal increments and treated as described above. Debt was valued each year by calculating the price of each portion

at the prevailing yield on industrials reported by Moody's. Maturities and coupons were taken into account.

The value of other liabilities was taken to be the book value of assets less the book values of common and preferred stock, debt, and current liabilities. Also, deferred taxes were subtracted out on the assumption that equity investors never expect to pay these non-interest-bearing amounts.

For the denominator of q, assets were categorized as plant and equipment, inventory, and other. Plant and equipment was valued by setting up an acquisition schedule and adjusting for price level and depreciation. Specifically, the value of plant and equipment at the start (1961) was assumed to be the book value. Each year, the value of plant and equipment was reduced by depreciation at 5 percent, then adjusted to the new price level according to the *GNP* implicit price deflator. New additions or sales were calculated as the change in gross plant at book value.

Inventory was adjusted for various accounting methods on the assumption that the firm used only that method reported for that firm by Compustat. The adjustment follows Lindenberg and Ross. For FIFO, any excess of beginning over ending inventory is price adjusted according to the wholesale price index, and any change in reported inventory is added at book. For LIFO, the beginning inventory is adjusted for a full year's inflation, and any change in reported inventory is adjusted for a half year's inflation. Average cost inventories are adjusted for one-half year's inflation. Inventories under the retail method are deflated according to the firm's average markup (i.e., sales divided by cost of goods sold). All other inventory accounting methods are assumed to result in equality of book and current value. Finally, other assets were assumed to have replacement cost equal to book value.

REFERENCES

Bain, Joe S., "Relation of Profit-Rate to Industry Concentration: American Manufacturing, 1936–1940," *Quarterly Journal of Economics*, August 1951, *65*, 293–324.

Belsley, David, Kuh, Edward and Welsch, Robert, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, New York: Wiley & Sons, 1980.

Brozen, Yale, "The Antitrust Task Force Deconcentration Recommendation," *Journal of Law and Economics*, October 1970, *13*, 279–92.

______, *Concentration, Mergers, and Public Policy*, New York: Macmillan, 1982.

Carter, John, "Collusion, Efficiency and Antitrust," *Journal of Law and Economics*, October 1978, *21*, 435–44.

Demsetz, Harold, "Industry Structure, Market Rivalry, and Public Policy," *Journal of Law and Economics*, April 1973, *16*, 1–9.

______, "Two Systems of Belief about Monopoly," in Harvey Goldschmid et al., eds., *Industrial Concentration: The New Learning*, Boston: Little, Brown & Co., 1974.

______, "More on Collusion and Advertising: A Reply," *Journal of Law and Economics*, April 1976, *19*, 204–08.

______, "Barriers to Entry," *American Economic Review*, March 1982, *72*, 47–57.

Edwards, Franklin, "Managerial Objectives in Regulated Industries: Expense-Preference Behavior in Banking," *Journal of Political Economy*, February 1977, *85*, 147–62.

Fisher, Franklin and McGowan, John, "On the Misuse of Accounting Rates of Return to Infer Monopoly Profits," *American Economic Review*, March 1983, *73*, 82–97.

Gale, Bradley, "Market Share and Rate of Return," *Review of Economics and Statistics*, November 1972, *54*, 412–23.

Goldschmid, Harvey, Mann, H. Michael and Weston, J. Fred, *Industrial Concentration: The New Learning*, Boston: Little, Brown & Co., 1974.

Hannan, Timothy, "Expense-Preference Behavior in Banking: A Reexamination," *Journal of Political Economy*, August 1979, *87*, 891–95.

Imel, Blake and Helmberger, Peter, "Estimation of Structure-Profits Relationships with Application to the Food-Processing Sector," *American Economic Review*, September 1971, *61*, 614–27.

Lindenberg, Eric and Ross, Stephen, "Tobin's q

Ratio and Industrial Organization," *Journal of Business*, January 1981, *54*, 1–32.

McGee, John, "Efficiency and Economies of Size," in Harvey Goldschmid et al., eds., *Industrial Concentration: The New Learning*, Boston: Little, Brown & Co., 1974.

Ornstein, Stanley, "Concentration and Profits," *Journal of Business*, October 1972, *45*, 519–41.

Peltzman, Sam, "The Gains and Losses from Industrial Concentration," *Journal of Law and Economics*, October 1977, *20*, 229–63.

Ravenscraft, David, "Collusion v. Superiority: A Monte Carlo Analysis," manuscript, Federal Trade Commission, 1981.

______, "Structure-Profit Relationships at the Line of Business and Industry Levels," *Review of Economics and Statistics*, February 1983, *65*, 22–32.

Schwert, G. William, "Using Financial Data to Measure Effects of Regulation," *Journal of Law and Economics*, March 1981, *24*, 121–58.

Shepherd, William, "The Elements of Market Structure," *Review of Economics and Statistics*, February 1972, *54*, 25–37.

Stigler, George, *Capital and Rates of Return in Manufacturing*, Princeton: Princeton University Press, 1963.

Sullivan, Timothy, "A Note on Market Power and Returns to Stockholders," *Review of Economics and Statistics*, February 1977, *59*, 108–13.

Thomadakis, Stavros B., "A Value-Based Test of Profitability and Market Structure," *Review of Economics and Statistics*, May 1977, *59*, 179–85.

Weiss, Leonard, "The Concentration-Profits Relationship and Antitrust," in Harvey Goldschmid et al., eds., *Industrial Concentration: The New Learning*, Boston: Brown, Little & Co., 1974.

______ **and Pascoe, George,** "Some Early Results of the Effects of Concentration from the FTC's Line of Business Data," manuscript, Federal Trade Commission, 1981.

The Degree of Fiscal Illusion in Interest Rates: Some Direct Estimates

By JOE PEEK AND JAMES A. WILCOX*

Michael Darby (1975), Martin Feldstein (1976), and Vito Tanzi (1976) have suggested that, in the presence of income taxation, nominal interest rates would have to change by more than expected inflation to preserve expected after-tax real interest rates. The frequent empirical rejection of this hypothesis has often been attributed to the failure of interest rates to allow for interest income taxation, a characteristic Tanzi (1980) has termed "fiscal illusion."[1] Here we investigate the degree of fiscal illusion by directly estimating the response of nominal interest rates to changes in tax rates. After deriving the reduced form for the interest rate in Section I, we demonstrate the difficulty in drawing inferences about fiscal illusion from existing estimates. In Section III, we present direct estimates of the extent of fiscal illusion.

I. The Model

The macro model we use is similar to that of our earlier article (1983), augmented with a fiscal illusion parameter. The *IS*, *LM*, wage, and aggregate supply relations (deflated by Y^N) can be expressed as

$$(1)\quad Y-Y^N=a_0-a_1r^*+a_2\Delta Y+a_3(X-Y^N)+a_4(M-P-Y^N)-a_5SS-a_6FB,$$

$$(2)\quad M-P-Y^N=b_0+b_1(Y-Y^N)-b_2i^*-b_3FB,$$

$$(3)\quad W=c_0+P^e-c_1SS,$$

$$(4)\quad P=d_0+W+d_1(Y-Y^N)+d_2SS,$$

where the coefficients of all the variables are positive, Y, Y^N, X, M, P, P^e, and W are real output, natural (i.e., potential) real output, the sum of real exports and real government expenditures, the nominal money supply, the price level, the expected price level, and the nominal wage, respectively (all in logs). ΔY is the percentage change in actual real output lagged one period, SS is a supply shock variable, FB is the domestic bonds held by foreigners, and r^* and i^* are tax-adjusted real and nominal interest rates. The nominal interest rate (i) is related to i^* and r^* by (5) and (6):

$$(5)\quad i^*\equiv(1-\theta t)i,$$

$$(6)\quad r^*\equiv i^*-p^e,$$

where t is the marginal tax rate on interest income, p^e is the anticipated inflation rate, and θ is the fiscal illusion parameter. A value of unity for θ implies that agents respond to after-tax, rather than pre-tax, interest rates and therefore do not suffer from fiscal illusion. A value for θ of zero, at the other end of the presumed range for this parameter, would imply that agents disregard taxes entirely, that is, suffer from complete fiscal illusion.

Real expenditures depend on the real interest rate after allowance for taxes and for the degree of fiscal illusion, real exogenous export and government demand, a real balance effect, and an investment accelerator term. The opportunity cost of holding money is the tax-adjusted nominal interest rate. The wage and price equations embody the natu-

*Department of Economics, Boston College, Chestnut Hill, MA 02167, and School of Business Administration, 350 Barrows Hall, University of California, Berkeley, CA 94720. We are grateful to Eugene Fama, Richard Meese, Jeffrey Skelton, Janet Yellen, and two anonymous referees for their comments. Josef Chytry provided accurate typing. Any errors are solely our responsibility. Financial support was provided by a Boston College Summer Research Fellowship and NSF grant SES-8109093.

[1] We feel that a more accurate term for this characteristic would be "tax illusion." Since it has already become widely known as "fiscal illusion," we will use this latter term.

ral rate hypothesis. A supply shock, for example a sudden increase in the relative price of imported oil, lowers the *IS* curve through its effect on the demand for capital, and hence investment demand (see Wilcox, 1983b), shifts the aggregate supply equation by raising the cost of production, and reduces the equilibrium real wage. The *FB* variable is included to isolate the financial effects arising from the supply shocks. In the *IS* curve, *FB* serves as a proxy for any increase in the world saving rate that developed as real income was transferred to countries (OPEC) with higher saving propensities. Similarly, *FB* enters the *LM* equation to allow for the possibility that the demand for money will be reduced as wealth is transferred to agents who desire a wealth portfolio with a much higher proportion of U.S. government securities than domestic wealth holders do.

Equations (1)–(6) yield the reduced-form equation for the after-tax nominal interest rate:

$$(7)\quad i=\frac{A_0}{(1-\theta t)}+\underset{(+)}{A_1\tilde{p}^e}+\underset{(?)}{A_2\tilde{M}'}+\underset{(+)}{A_3\tilde{X}'}+\underset{(+)}{A_4\Delta\tilde{Y}}+\underset{(?)}{A_5\widetilde{SS}}+\underset{(-)}{A_6\widetilde{FB}},$$

where M' and X' are $(M-P^e-Y^N)$ and $(X-Y^N)$, respectively, and a tilde over a variable indicates that it has been divided by $(1-\theta t)$. The liquidity and real balance effects of an increase in the real money supply have offsetting effects on the interest rate resulting in an ambiguous sign for A_2. Likewise, the sign of A_5 is indeterminate a priori.[2]

The reduced-form effect of expected inflation on interest rates is

$$(8)\quad di/dp^e=\beta=A_1/(1-\theta t)$$
$$=1\Big/\left[(1-\theta t)1+\left\{\frac{b_2(1+a_4d_1)}{a_1(b_1+d_1)}\right\}\right]$$

Equation (8) highlights the fact that very little can be deduced about the presence or strength of tax effects on interest rates from estimates of β.[3] To test for the presence of tax effects, we disentangle the tax $(1-\theta t)$ and nontax (A_1) terms that comprise β. We take the Darby hypothesis to be that $\beta=1/(1-t)$. Equivalently, this is the joint hypothesis that $A_1=1$ and $\theta=1$. We distinguish this from the simple hypothesis that $\theta=1$, the "complete tax adjustment" or "absence of fiscal illusion" hypothesis. This hypothesis asserts only that individuals base their behavior on after-tax interest rates. An even less restrictive version of this hypothesis would be that tax rates are not completely ignored: $\theta>0$.

Clearly, the existence of fiscal illusion ($\theta<1$) would imply rejection of the Darby hypothesis. Rejection of the joint Darby hypothesis, however, does not necessarily imply rejection of the hypothesis of complete tax adjustment ($\theta=1$). Nor does the finding that $\theta=1$ imply that the Darby hypothesis holds, since A_1 will be less than unity unless, for example, either the *IS* curve is horizontal or the *LM* curve is vertical (see equation (8)).

II. Previous Tests for Tax Effects

Early tests of the Darby hypothesis (see, for example, Thomas Cargill, 1977; John Carlson, 1979; Jack Carr, James Pesando, and Lawrence Smith, 1976) estimated reduced-form nominal interest rate equations where tax effects remained embedded in the reduced-form coefficient, β. That hypothesis did not receive much empirical support.[4] Tanzi (1980) did separate β into its tax and nontax components, employing $p^e/(1-t)$ as an explanatory variable. He rejected the hypothesis that the coefficient on this variable

[2] The investment-real wage effects of supply shocks might be expected to dominate, suggesting a negative value for A_5. The results presented in Wilcox (1983a, b) and our earlier paper can be so interpreted.

[3] This has been pointed out by Maurice Levi and John Makin (1978), Michael Melvin (1982), and Peek (1982).

[4] Cargill and Robert Meyer (1980) found significantly greater than unity estimates for the 1960's, but they disappeared when the sample period was extended into the 1970's. Wilcox (1983b) suggested a cause for this coefficient decline.

(our A_1) was unity and concluded that people suffered from fiscal illusion. From (8), however, we see that this procedure assumes $\theta = 1$ (i.e., the absence of fiscal illusion) and tests whether A_1 differs from unity, thereby precluding estimates of, tests for, or conclusions about the extent to which interest rates react to changes in tax rates.

Michael Melvin recognizes this difficulty. Using structural parameter estimates for a small macro model similar to the one presented above in Section I, he obtains a value of 0.511 for A_1. This estimate, however, is conditioned on $\theta = 1$. Although Tanzi and Melvin obtain similar estimates, Melvin correctly notes that a coefficient below unity does not imply that taxes are ignored. On the other hand, his estimates do not imply that there *are* tax effects.

In a recent article, Robert Ayanian proposes an alternative test of the Darby hypothesis. His regression of the taxable on the tax-exempt yield provides a coefficient of 1.63 ($s.e. = 0.04$). He interprets this as an estimate of $1/(1-t)$ and concludes that there is "an unmistakable Darby effect" (1983, p. 763). Although the spread between taxable and tax-exempt yields almost certainly reflects tax rates (and is, in fact, often used to obtain a proxy for them—see Darby), Ayanian's results do not indicate "whether or not the expected real rate was depressed by expectations of inflation" (p. 764), whether the taxable rate rises with taxes, whether the tax-exempt rate is invariant to tax rate changes, or whether the spread has responded to tax rate changes.

Using Russell Davidson and James MacKinnon's (1981) model specification tests, Peek and Peek-Wilcox were able to distinguish between the tax-adjusted and non-tax-adjusted Fisher hypotheses. Faced with a choice between the two extremes of complete tax adjustment ($\theta = 1$) and complete fiscal illusion ($\theta = 0$), these tests implied the rejection of complete fiscal illusion and failure to reject the complete tax adjustment hypothesis. Although these conclusions are not based on θ-conditioned estimates of A_1, they provide no estimate of the *degree* of fiscal illusion.

III. Direct Estimates of the Degree of Fiscal Illusion

We obtain a direct estimate of the degree of fiscal illusion, θ, by nonlinear least squares estimation of (7). Table 1 presents the results. The estimates are based on semiannual monthly observations (June and December) to match the Livingston survey data. The sample extends from June 1952 through June 1979. This sample period avoids the pre-1952 pegging of interest rates by the Federal Reserve, the imposition of credit controls in 1980, and any structural changes associated with financial deregulation and monetary policy after June 1979. Monthly averages of the one-year Treasury bill bond-equivalent yield during June and December are used as the before-tax nominal interest rate measure (i) to match the maturity of the Livingston one-year anticipated inflation rate data.

The anticipated inflation rate series, PE, is the percentage change in the CPI expected over the next twelve months derived from the Livingston survey. This series was provided by the Federal Reserve Bank of Philadelphia. This measure of anticipated inflation has two advantages: it is a truly *ex ante* expectation and reflects whatever sophistication agents use to process information.

Second- and fourth-quarter observations are used for the remaining explanatory variables. The logarithm of the sum of real exports and real government expenditures on goods and services divided by the level of natural real output (X') and the percentage change in real GNP lagged one period (ΔY) are constructed from the *National Income and Product Accounts* data. We use the potential real GNP series constructed by the Council of Economic Advisors as our measure of natural real output. The logarithm of the nominal money supply deflated by the expected price level and natural real output (M') is constructed using the $M1$ definition of the money supply and the Livingston survey measure of the expected price level. The tax rate (t), the supply shock variable (SS), and the foreign holdings of bonds (FB) are described in detail in our earlier

TABLE 1—REDUCED-FORM ESTIMATES FOR INTEREST RATES 1952:06–1979:06, SEMIANNUAL OBSERVATIONS

	Dependent Variable				
	i (1)	i (2)	i (3)	i (4)	i_x (5)
Constant	14.2	12.2	2.4	5.7	12.4
	(4.92)	(3.81)	(1.25)	(1.93)	(2.22)
PE	0.821	0.654	–	–	0.681
	(7.23)	(3.47)			(1.99)
$PEIN$	–	–	0.330	–	–
			(4.20)		
$PEOUT$	–	–	–	0.122	–
				(1.83)	
M'	1.94	1.48	−1.12	−3.91	2.06
	(1.62)	(1.36)	(1.86)	(2.95)	(1.32)
X'	3.93	3.37	0.75	5.06	3.32
	(2.16)	(2.05)	(0.51)	(2.05)	(1.52)
ΔY	7.00	5.47	3.64	5.33	2.51
	(2.65)	(2.05)	(1.70)	(1.43)	(0.88)
SS	−3.22	−2.94	−1.26	−2.64	−2.55
	(5.99)	(5.55)	(3.08)	(3.84)	(2.88)
FB	−4.68	−4.03	−1.10	−2.03	−4.84
	(2.55)	(2.47)	(0.80)	(1.03)	(2.02)
θ	1.00	1.40	1.34	1.33	0.26
		(3.52)	(3.77)	(2.21)	(0.19)
R^2	.909	.910	.931	.856	.851
$D\text{-}W$	1.64	1.66	2.21	1.39	1.82
SEE	0.730	0.733	0.639	0.925	0.525

Notes: Data and sources are given in text; absolute values of t-statistics are shown in parentheses.

paper. The tax rate is calculated as a weighted average of the marginal personal income tax rate for each adjusted gross income class.[5] The variable SS is measured by the ratio of the implicit price deflator for imports to the GNP deflator adjusted for exchange rate changes. FB is the ratio of foreign holdings to the sum of private domestic and foreign holdings of U.S. government short-term marketable securities.

In the table, column (1) is obtained when (7) is estimated under the constraint that interest rates adjust completely to changes in tax rates ($\theta = 1$). The estimation method is ordinary least squares. Expected inflation, exogenous expenditures, the change in real income, supply shocks, and foreign demand for bonds each enter significantly. Column (2) allows for a freely estimated fiscal illusion term, θ. The point estimate of 1.40 is insignificantly different from one. This estimate suggests that the adjustment to tax rate changes is complete: pre-tax interest rates rise by enough to preserve after-tax yields. Further, we can easily reject the hypothesis of complete fiscal illusion ($\theta = 0$).[6]

Columns (3) and (4) substitute alternative expected inflation measures, $PEIN$ and $PEOUT$, for the Livingston survey measure, PE. The $PEIN$ and $PEOUT$ are in-sample and out-of-sample forecasts based on infor-

[5]For the June observations, we use the tax rate for that calendar year. For the December observations, we use the average of the tax rates for the current year and for the upcoming year.

[6]These point estimates differ somewhat from our earlier paper due to minor data revisions and because here the dependent variable is calculated as the bond-equivalent yield as opposed to a discount factor. The results are not sensitive to the choice of yields. When the interest rate is calculated on a discount basis, the estimated value of θ is 1.23 ($t = 2.65$).

mation contained in earlier Treasury bill yields.[7] Using monthly data, inflation (twelve times the month over month change in the log of the *CPI*) is regressed on a constant and six lags of the one-month Treasury bill yield. *PEIN* is the vector of fitted values for June and December obtained using the entire 1952:06–1979:06 sample. The out-of-sample forecasts, *PEOUT*, are based only on prior information. Thus, the forecast during June depends on the six monthly-average Treasury yields from December to May and the forecast equation coefficients. These coefficients are obtained by regressing inflation on a constant and six lags of one-month Treasury yields over the forty-eight months ending two months before the forecast is made. Using coefficient estimates from a sample that edged closer to the forecast dates (June and December) would require more information than agents actually had. Most of the coefficients, especially those on the expected inflation measures themselves, are affected by this substitution. The estimates of the tax-adjustment parameter θ, however, are virtually unchanged. Thus, the finding that interest rates respond completely to changes in tax rates is robust with respect to the measure of expected inflation.

Column (5) replaces the taxable Treasury bill yield with the one-year tax-exempt municipal bond yield, i_x, obtained from Salomon Brothers' *Analytical Record of Yields and Yield Spreads*. If after-tax yields are tax-invariant as indicated by column (2), we would expect tax-exempt yields to not respond to tax rate changes. The estimate for θ of -0.26, which is insignificantly different from zero, implies that those yields are hardly affected by taxes. Since our model does not indicate the appropriate specification for testing the effects of tax rates on tax-exempt yields, two additional forms were also estimated. Setting θ equal to zero, we reestimated column (5), once adding t and once adding $1/(1-t)$ as explanatory variables. Their coefficients were 1.17 and 0.69 with t-statistics of 0.24 and 0.27, respectively. Thus changes in tax rates leave after-tax yields unaffected.

[7]Eugene Fama and Michael Gibbons (1984) argue that forecasts based on interest rates may be superior to survey forecasts.

IV. Conclusion

We have argued that previous studies have not produced tests that permit inference about either the presence or the degree of fiscal illusion in interest rate determination. The specification we use allows us to estimate directly the extent to which interest rates adjust to changes in tax rates. The estimates based on taxable and on tax-exempt yields and on various measures of expected inflation imply that after-tax real yields are invariant with respect to tax rate changes. Thus, these results do not suggest there has been fiscal illusion.

REFERENCES

Ayanian, Robert, "Expectations, Taxes, and Interest: The Search for the Darby Effect," *American Economic Review*, September 1983, *73*, 762–65.

Cargill, Thomas F., "Direct Evidence of the Darby Hypothesis for the United States," *Economic Inquiry*, January 1977, *15*, 132–34.

______ **and Meyer, Robert A.,** "The Term Structure of Inflationary Expectations and Market Efficiency," *Journal of Finance*, March 1980, *35*, 57–70.

Carlson, John A., "Expected Inflation and Interest Rates," *Economic Inquiry*, October 1979, *17*, 597–608.

Carr, Jack, Pesando, James E. and Smith, Lawrence B., "Tax Effects, Price Expectations and the Nominal Rate of Interest," *Economic Inquiry*, June 1976, *14*, 259–69.

Darby, Michael R., "The Financial and Tax Effects of Monetary Policy on Interest Rates," *Economic Inquiry*, June 1975, *14*, 266–76.

Davidson, Russell and MacKinnon, James G., "Several Tests for Model Specification in the Presence of Alternative Hypotheses," *Econometrica*, May 1981, *49*, 781–93.

Fama, Eugene F. and Gibbons, Michael R., "A Comparison of Inflation Forecasts," *Jour-*

nal of Monetary Economics, May 1984, *13*, 327–48.

Feldstein, Martin S., "Inflation, Income Taxes, and the Rate of Interest: A Theoretical Analysis," *American Economic Review*, December 1976, *66*, 809–20.

Levi, Maurice D. and Makin, John H., "Anticipated Inflation and Interest Rates: Further Interpretation of Findings on the Fisher Equation," *American Economic Review*, December 1978, *68*, 801–12.

Melvin, Michael, "Expected Inflation, Taxation, and Interest Rates: The Delusion of Fiscal Illusion," *American Economic Review*, September 1982, *72*, 841–45.

Peek, Joe, "Interest Rates, Income Taxes, and Anticipated Inflation," *American Economic Review*, December 1982, *72*, 980–91.

______ **and Wilcox, James A.**, "The Postwar Stability of the Fisher Effect," *Journal of Finance*, September 1983, *38*, 1111–24.

Tanzi, Vito, "Inflation, Indexation and Interest Income Taxation," *Banca Nazionale del Lavoro-Quarterly Review*, March 1976, No. 116, 54–76.

______, "Inflationary Expectations, Economic Activity, Taxes, and Interest Rates," *American Economic Review*, March 1980, *70*, 12–21.

Wilcox, James A., (1983a) "The Effects of Inflation Uncertainty and Supply Shocks on Real Interest Rates," *Economics Letters*, No. 2, 1983, *12*, 163–67.

______, (1983b) "Why Real Interest Rates Were So Low in the 1970s," *American Economic Review*, March 1983, *73*, 44–53.

Salomon Brothers Inc., *Analytical Record of Yields and Yield Spreads*, New York.

U.S. Department of Commerce, Bureau of Economic Analysis, *The National Income and Product Accounts of the United States, 1929–76 Statistical Tables*, Washington: USGPO, September 1981.

______, *Survey of Current Business*, Washington: USGPO, various issues through July 1982.

The Greek Stabilization of 1944–46

By Gail E. Makinen*

It is a matter of great contention among economists whether economic stabilization programs can be instituted without imposing high real costs. Those believing in the core or underlying rate of inflation hypothesis argue that relying only on conventional monetary and fiscal policies will impose high costs. The rational expectations proponents, on the other hand, argue that a convincing anti-inflation program will likely minimize these costs as economic agents respond to a genuine regime change or change in the rules under which monetary and fiscal policies are conducted.[1] A way of discriminating between these contending views is to examine the stabilization phase of the world's episodes of hyperinflation. Thus far, the work of Thomas Sargent (1982) on the post-World War I experiences in Austria, Hungary, Poland, and Germany, and William Bomberger and myself (1983) on post-World War II Hungary have adduced evidence in support of the rationalists' view. Price level stability was achieved rapidly without a prolonged period of high unemployment. The Greek stabilization of 1944–46 is not so straightforward and, as such, provides an interesting contrast to the other episodes. Its' unique feature is that price level stability took over a year to achieve following the initial reform of November 11, 1944. It was not achieved until, in a third reform in early 1946, the Greek

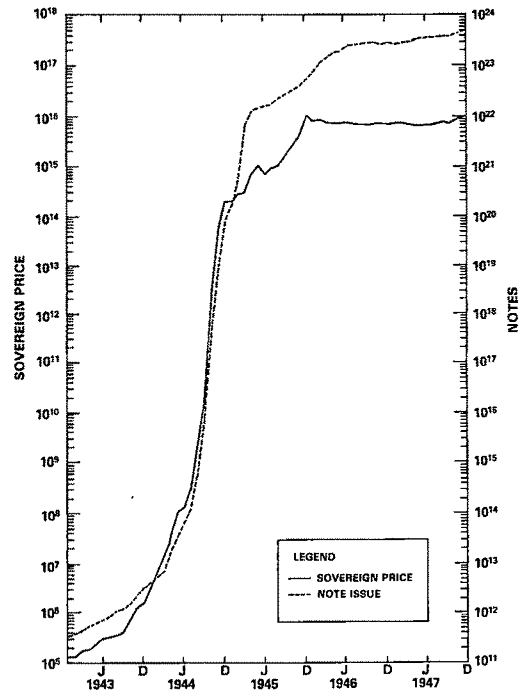


Figure 1. Price of the Gold Sovereign and Note Issue (End of Month Data)

government put in place those features identified by the other studies as essential to the successful stabilization efforts they examined. Price level stability followed immediately in an environment where unemployment apparently declined. The Greek episode thus provides a good setting for evaluating the alternative views on stabilization as well as elements of successful and unsuccessful stabilizations.

To give some perspective to the period to be studied, in Figure 1 data on the money stock (measured as notes of the Bank of Greece) and the market price of the gold sovereign are recorded for the period 1943–48; a period both before the hyperinflation and after the successful stabilization. (Source

*Specialist in Economic Policy, Congressional Research Service, Library of Congress, Washington, D.C. 20540. The views expressed in this paper are my own and do not represent the views of the Congressional Research Service or the Library of Congress. This study has benefited from constructive comments by Phillip Cagan, Gardiner Patterson, Thomas Woodward, Robert Anderson, William Bomberger, and two anonymous referees. My understanding of the stabilization effort was substantially enhanced by reading Patterson's dissertation.

[1] The core or underlying inflation view of stabilization is best identified with Arthur Okun (1978), James Tobin (1980), and Otto Eckstein (1981).

for Figures 1 and 2 is D. Delivanis and W. Cleveland, 1949.) The market price of the sovereign was selected to reflect prices in general for two reasons. First, no price index exists on a monthly basis which covers the 1943–48 period. Second, it is the least encumbered by prices which were either frozen (rents) or set by government decree (as were many food prices).

I. Stabilization

While the particulars of the various stabilization efforts differ, the prior studies identified two specific measures as essential to their success. First, an independent central bank was reconstituted with the power to refuse the government's demand for unsecured credit. Second, fiscal policy was substantially overhauled so that through a combination of tax increases and expenditure reductions, the government budget in the near term and over the longer run was brought into reasonable balance (if not surplus).

To the extent that these changes are convincing, rational economic agents will perceive a regime change and revise downward their expectations of future inflation, and a phenomenon associated with successful stabilizations will be observed: a substantial increase in high-powered money occurs without a parallel movement in prices (i.e., real money balances rise). Supposedly, these same forecasts of inflation should lead to a moderation in wage demands so that unemployment need not rise during the stabilization.[2]

The stabilization phase of the Greek hyperinflation will be discussed in terms of the initial event which occurred on November 11, 1944, the second attempt, the Varvaressos Reform, in June 1945, and, finally, the third effort, the Anglo-Hellenic Convention concluded on January 24, 1946. This will be followed by an analysis of the behavior of unemployment and real balances.

A. *November 11, 1944*

This initial effort at reform had but two essential features. The first limited the government to an overdraft of 2 billion drachma at the Bank of Greece. The second involved a massive conversion of old drachma for new at a rate of 50 billion to one.[3] The new drachma was then made convertible into British Miliary Authority (BMA) pounds at the rate of 600 to one (but only in lots larger than 12 thousand drachma).[4] Old and new drachma and BMA pounds were made legal tender, although the latter was not legal tender in Britain. No changes were made in the tax system nor were expenditure cuts announced. However, Greece expected to be the recipient of massive amounts of foreign aid that was to be sold and the revenue used for budget purposes.[5]

The results of this first stabilization effort were less than promising. Within the first two months the government had exceeded its

[2] In his study, Sargent notes that the foreign exchange rates also stabilized as well. Data on these rates were not available to Bomberger and me.

[3] By way of contrast, only in Germany and Hungary II did the initial reform coincide with the introduction of a new unit of account. In Poland the introduction was delayed several months and in the cases of Austria and Hungary I by several years. In all of the other episodes the reform legislation forbade the governments from borrowing from the central bank on an unsecured basis. However, in the cases of Germany, Poland, and Hungary II, the governments were allowed an initial limited overdraft in order to finance expenditures until the new taxes yielded revenue. Finally, in both Germany and Hungary II, an absolute ceiling was initially placed on the amount of notes the central bank could issue.

[4] This convertibility placed no effective limit on the fiscal activities of the Greek government. By secret treaty, the British had agreed to advance a maximum of 3 million BMA pounds. This ceiling was never reached (see Gardiner Patterson, 1948, pp. 58–60). In essence, this convertibility was analogous to the relationship of the U.S. Treasury to the Federal Reserve in determining the composition of the U.S. money supply.

[5] At the time of the reform, the British were furnishing substantial aid to the Greeks. On April 1, 1945, the aid program was taken over by UNRRA. From April to December 1945, UNRRA spent $350 million which, if one disregards price level changes, was equal to about one-half of the 1938 national income of Greece. No other stabilization effort benefited from this type of aid. In Austria and Hungary I, stabilization was supported by foreign loans and, in the case of Germany, the Dawes Commission suspended her reparations payments and rescheduled the remainder so to reduce their burden.

TABLE 1—SELECTED ECONOMIC DATA, DECEMBER 1944–DECEMBER 1945

Percentages:	Dec.	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Rate of Inflation[a]	63.2	−8.1	5.3	6.7	1.6	40.0	−38.5	5.4	69.5	55.0	51.6	42.1	126.3
Rate of Note Issue[a]	990.0	270.0	78.0	42.0	65.0	22.0	17.4	9.2	22.4	15.9	31.3	25.1	31.3
Rate of Rise in the Price of Sovereigns	47.6	35.5	7.1	20.0	122.2	58.3	−36.8	33.3	12.5	66.7	46.6	59.1	160.0
Expenditures Covered by Receipts		2.3[b]	6.5	30.5	36.4	51.7	47.2	75.0	57.5	52.3	50.9	55.5	37.7

Source: Delivanis and Cleveland (p. 188).
[a]Percentage for December calculated from November 11, 1944.
[b]The 2.3 percent represents total for months of December 1944 and January 1945.

legal overdraft at the Bank of Greece and thereafter operated in disregard of the law. During the first seven months after stabilization, prices, as shown by the cost of living index in Table 1, rose 140 percent (measured as May over November). Since this index contains commodities and services (principally rent) fixed or controlled by government edict, it likely understates price movements in general. As a substitute, if the price of the gold sovereign, the principal alternative to commodities as an inflation hedge, is used, the rise is over 800 percent.[6] Taxes yielded less than one-sixth of total revenue and, while the sale of aid goods yielded almost 45 percent of total revenue, their distribution costs were high. So high, in fact, that for all of 1945, the aid program yielded no net revenue (gross revenue was 16.4 billion drachma while distribution costs were 16.8 billion).

The failure of the economy to revive and the acceleration of the inflation rate in April and May convinced the Greek government that a new reform was needed. On June 3, Kyriakos Varvaressos, a prominent economist, was named as a sort of economic czar.

B. *The Varvaressos Reform of June 4, 1945*

As Varvaressos saw it, the failure of the Greek economy to recover was due to a lack of effective state control over the economy and inadequate assistance from abroad. To remedy affairs, the government: 1) devalued the drachma (the official rate was still under the black market rate); 2) accelerated the arrival and distribution of foreign aid; 3) raised wages, especially at the lower level; 4) imposed wage and price controls; and 5) reduced by one-half the prices at which aid goods were sold. As provisions 3, 4, and 5 adversely affected the Greek budget, a large tax was imposed on the occupants of rented dwellings (rents having been frozen in Greece since 1940).

Varvaressos did not succeed largely because he antagonized a politically powerful group with his rent tax and the government could not guarantee supplies of those items subject to price controls. He resigned on September 1.

During his three months in office, prices fell in June, rose slightly in July, and sharply in August (see Table 1). The gold sovereign price shows a similar movement. State finances recorded some improvement. Once the potency of the rent tax was reduced, the fiscal situation deteriorated during the remainder of 1945 and inflation accelerated.

In December, the British initiated a major effort to stabilize the Greek economy. The result was the Anglo-Hellenic Convention of 1946.

C. *The Anglo-Hellenic Convention, January 24, 1946*

The provisions of this convention, embodying the reforms common to the other

[6]The foreign exchange rates did not stabilize, the black market rates continued to rise, and the gap between them and the official rates widened over this period.

TABLE 2—SELECTED ECONOMIC DATA FOR 1946

	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Rate of Inflation[a]	32.7	−13.6	2.1	2.1	1.3	0.2	−0.2	−3.1	3.2	3.1	3.7	−5.7
Rate of Note Issue	33.6	61.6	27.5	30.4	7.1	5.8	7.8	11.6	3.2	−1.2	−7.4	14.8
Rate of Change in the Price of Sovereigns	−17.6	−4.5	−4.1	–	−0.9	1.7	1.7	–	0.5	–	0.4	–

Source: See Table 1.
[a]The inflation rate for January is computed using the old price index. Data for subsequent months using this index are unavailable.

successful episodes, contributed substantially to a final and successful stabilization.

First, British experts helped to revamp the fiscal system by (a) preparing a realistic budget, (b) improving the tax assessment and collection administration, (c) adjusting many specific taxes for past inflation, and (d) increasing substantially the prices at which aid goods were sold. Had it not been for the outbreak of civil war in September 1946, the budget would have been balanced.

Second, while an independent central bank was not reconstituted, a five-member Currency Committee was created whose unanimous consent was required before the Bank of Greece could issue any notes. The Committee came to exercise a pervasive control over the Greek economy and public finances.[7]

Third, to enforce fiscal and monetary austerity, the Bank of Greece pledged to maintain a fixed exchange rate vis-à-vis the dollar and pound sterling through open market operations in gold sovereigns. This restored the gold convertibility of the drachma for Greek citizens.[8]

Fourth, to provide the financial wherewithal to carryout this program, the British cancelled their loan of £46.5 million made to Greece in 1940–41. This unencumbered over one-half of the reserves of the Bank of Greece. In addition, they made a loan of £10.0 million to serve as a cover for the note issue.[9] The United States furnished credits of $80 million dollars.[10]

The effect of policy based on the Convention was immediate. The Bank of Greece commenced open market sales of gold sovereigns and maintained the drachma-pound-dollar exchange rate. The price level as shown in Table 2 not only stabilized but a mild deflation set in. While state receipts covered only 45 and 36.5 percent of expenditures in January and February, they rose to over 50 percent in March. The fiscal picture improved markedly thereafter. This is demonstrated by the fact that whereas the government borrowed 204 billion drachma from the Bank of Greece for the first two months after the Convention, it only required 196 billion for the final nine months of 1946.

II. Evidence Bearing on a Regime Change

A. *Unemployment and Output*

Any interpretation of the unemployment and output effects of stabilization in each of

[7]The Committee contained an American and a British member. While it abridged the sovereignity of Greece, both Austria and Hungary I had to admit a League of Nations supervisor to ensure that the provisions of their stabiliziation programs were being carried out.

[8]At this time, the drachma was devalued and set at rates closely approximating those on the blackmarket. Greece was then one of the only countries in the world with an internal gold standard. In no other stabilization program was gold convertibility restored.

[9]This reserve was more superficial than substantive since these notes had no gold reserve requirements or other specific backing. This is not true of the other stabilizations.

[10]Since the British were very instrumental in stabilizing the economies of Austria and Hungary at the end of World War I and, at the time the Committee was functioning, Hungary was busy stabilizing its own monetary system, I was interested to learn the ways these experiences were being used in Greece. I was told that the British may have used them in their plans made in London, but they were never mentioned in the deliberations of the Currency Committee (Patterson, 1983).

the episodes is clouded by individual real dislocations or shocks that each economy suffered. After World War I, Austria and Hungary emerged as independent countries greatly reduced in size with the need to realign their economies accordingly. Poland was newly created and Germany lost substantial territory in the east and west. Both Hungary and Greece were major battlefields during World War II. Greece, in addition, had to readjust her foreign trade as her major prewar trading partners were Axis countries and those in the newly emerging Soviet sphere of influence.

The data assembled by Sargent show that while unemployment rose substantially in Austria, it was not as serious a problem in Hungary; that while it rose substantially in Poland it was no worse than in some months before stabilization; and that in Germany unemployment decreased.[11] Bomberger and myself found that unemployment rose in Hungary II, but not by a magnitude or duration that would be suggested by those adhering to the concept of a core rate of inflation.

Unfortunately, no data on unemployment were ever collected on a systematic basis at this time in Greece. Two observations do exist, however, for the period following the Anglo-Hellenic Convention. A survey undertaken in the summer of 1946 fixed the Greek population at 7.257 million and the labor force (14 years and older) at 2.663 million, of whom 197,000 or 7.4 percent were unemployed.[12] On December 31, 1947, the Greek Ministry of Labor reported unemployment at 122,000.[13] Whether the two methods of estimating the unemployed are comparable is unknown. Reinforcing the conclusion that unemployment fell are data on the real national income of Greece: it rose 5.5, 62.1, and 33.9 percent, respectively, in 1945, 1946, and 1947.[14]

Because of the real dislocations surrounding the economies during each of the stabilization episodes, the employment and output results may only be circumstantial evidence bearing on rational behavior. For that reason, a study of the movements of real balances is of interest.

B. *Real Balance Holdings*

If one starts from the premise that economic agents in Greece realized that the hyperinflation of 1943-44 was caused by a chronically unbalanced budget covered by recourse to advances from the Bank of Greece, one must then ask whether the set of events on November 11, 1944, convinced them that the fiscal money supply regime had changed such that they could expect future stability in the price level.

For economic agents to believe that the November 11 reform was a regime change they had to believe that 1) convertibility of the drachma into BMA pounds, 2) the governments 2 billion drachma overdraft at the Bank of Greece, and 3) the promise of substantial foreign aid would be sufficient to limit new money issues to noninflationary proportions. Does the movement in real money balances show us they were convinced?

In Figure 2, the log of real money balances is plotted for the twenty-five months following this first effort at reform. While three different prices indices are used to deflate the note series, all tell a similar story. The analysis will, however, use the series deflated by the market price of the gold sovereign for, as noted earlier, it is the least encumbered by prices that were either frozen or subject to state control.

In examining these data, one notices immediately that they do not exhibit the more

[11] The reference for Sargent's conclusion on Germany is F. D. Graham (1930, p. 287). However, Graham reports (p. 317) that among trade union members, those wholly unemployed rose from 6.3 percent in August 1923 to 28.2 percent in December (when the series ends) while those partially unemployed rose from 26.0 percent in August to 42.0 percent in December 1923. Thus, while unemployment initially rose dramatically, it was short-lived and did not persist into 1924.

[12] This survey was undertaken by the Allied Mission for Observing the Greek Election and is analyzed in Raymond Jessen et al. (1947).

[13] See Bank of Greece (1948, p. 72).

[14] National income data for 1945-47 are reported in High Board of Reconstruction (1950); for 1944, see Delivanis and Cleveland.

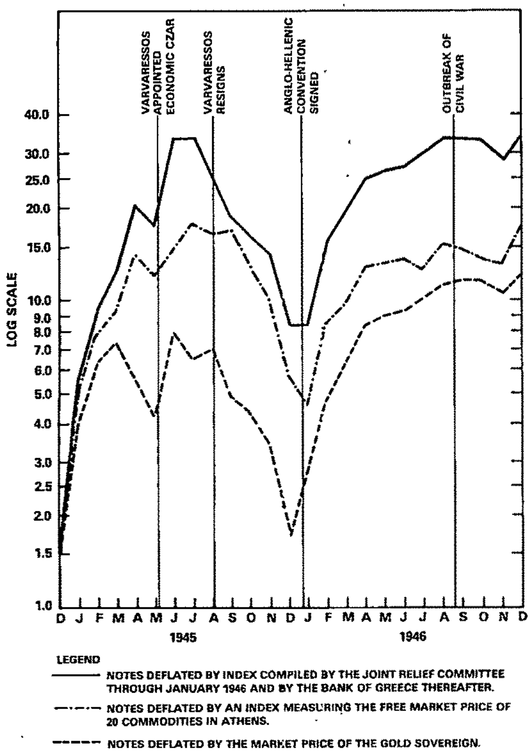


FIGURE 2. LOG OF REAL MONEY BALANCES (End of Month Data)

or less uniform increase over time that characterizes similar data in the other studies. Rather, for 1945, the level of real balances shows large fluctuations—a rapid rise followed by an equally spectacular decline. Is this pattern, so much at variance with the other studies, consistent with rational expectations?

To answer this question, two somewhat different models of rational behavior should be distinguished. One implies that the economy is inhabited by a sort of omniscient agent who, when presented with a new set of information, is able to assess its significance quickly and act accordingly. The agent is able to do so even in those instances in which there could be a good deal of uncertainty about important pieces of information; an uncertainty that can only be reduced by observing the government's actual performance. This model is a kind of ultrarational expectations view of the world. The other behavior model implies that agents are less omniscient. In many instances they can be expected to consume more time in gathering and processing the significance of new information. The additional time may be due to the uncertainty of the information. This should not be taken to mean that expectations are formed adaptively. Agents are still assumed to be forming expectations based on information concerning the monetary and fiscal policies to be pursued by the government.

If the ominscient behavior model is applicable, one might have expected the Greeks to have rejected the November 11 reform as being without substance, as it did not contain the crucial elements identified by past studies as being important in stabilizing economies—particularly, the overhaul of the fiscal system and the creation of an independent monetary authority. Nevertheless, the fivefold increase in real note balances from December through March 1945 far exceeds similar behavior in the other stabilization efforts.[15] Obviously, the Greeks reacted positively to this reform. How might this behavior be explained? The explanation makes use of the less than omniscient behavior model in a world in which policy announcements are clouded with uncertainty.

Using this model, the explanation holds that agents had some notion that the November 11 reform indicated that the government was serious in trying to deal with the economic crisis. However, one major uncertainty in their evaluation of the reform's substance was the likely value of foreign aid to the Greek budget, this being the crucial element in the Greek effort to achieve a balanced budget. While the dollar value of this aid was known to be substantial, its value to the Greek budget was highly uncertain. To reduce this uncertainty required information on how much aid was to be delivered, when it was to be delivered, the prices at which it was to be sold, and the

[15] Comparable increases were twofold for Austria (Sept.–Dec. 1922), 1.7-fold for Hungary (Mar.–June 1924), and 2.4-fold (Aug.–Oct. 1946), 1.5-fold for Germany (Dec.–Mar. 1924) and 1.8-fold for Poland (Jan.–Apr. 1924).

costs to distribute it. Only the passage of time in which the government's performance could be observed would yield information to reduce the uncertainty on these crucial elements and enable the Greeks to see whether the reform had substance. The evidence presented suggests that it did not take long for information to become available that the net revenue from this aid would be small. Since this outcome was obvious, the Greeks must then have entertained little hope for a balanced budget and noninflationary issues of money, that is, the reform was seen to be without substance. As a consequence, real balances declined in April and May of 1945.

This decline was interrupted by the appointment of Varvaressos as economic czar with his startlingly new rent tax, whose potential to raise revenue was great. Since the elements of his reform program were contradictory, time was again required to assess their substance. Once he resigned and the potency of the rent tax was reduced, economic agents became convinced that this reform was also without substance. The entire preceding nine months were then viewed as having consisted of two temporary changes in policy in the context of a regime remarkably like the one that had generated the hyperinflation. As a result, real money balances declined precipitously, and, in December 1945, they were barely above their level one year earlier. Whether they would have continued to decline cannot be known for Greece and Great Britian signed the Convention in January 1946.

Once the Anglo-Hellenic Convention was signed and the institutional changes put in place, the actual money supply process did in fact change. While it remained true that the budget deficit would still require borrowing from the Bank of Greece, the resultant note issue would be held to noninflationary proportions through open market operations in gold sovereigns. Under these circumstances, did individuals perceive that this reform had substance (i.e., that a regime change had taken place)? The evidence suggests that they did and did so quickly.

We observe behavior much like that noted in the other studies. The fourfold rise in real note balances during 1946 closely approximates in magnitude similar behavior one year after the German and second Hungarian efforts were undertaken. In addition, this rise in real balances is more or less continuous, which is also similar to all of the other European stabilization experiences.[16]

III. Conclusion

Economic behavior in Greece ultimately paralleled that observed in the other stabilization episodes. It took three reforms and success was not achieved until the Greeks put in place those changes the other studies identified as crucial to a regime change. The Greek experience, however, somewhat weakens the case for ultrarational expectations. The behavior of real balances during the first two reforms suggests that the public could not fully evaluate the possibility that these reforms constituted genuine regime changes without additional time to evaluate the government's actual performance—even when the nature of the reform might have suggested that a genuine regime change had not taken place.

One can conclude that, while the adaptive expectations-unit cost-core inflation view of the world is deficient as an explanation of Greek behavior during stabilization, the highly idealized version of rational expectations may not be entirely consistent with the evidence either. Rather, it may take time for individuals to evaluate and process new information.

[16] One year after Austria stabilized, real note balances had increased 2.6-fold, in Hungary 2.6-fold (first hyperinflation) and 4.4-fold (second hyperinflation), Germany 3.8-fold, and Poland 1.8-fold.

REFERENCES

Bomberger, William A. and Makinen, Gail E., "The Hungarian Hyperinflation and Stabilization of 1945–46," *Journal of Political Economy*, October 1983, *91*, 801–24.

Delivanis, D. and Cleveland, W., *Greek Mone-*

tary Developments 1939–1948 Bloomington: Indiana University Press, 1949.

Eckstein, Otto, *Core Inflation*, Englewood Cliffs: Prentice Hall, 1981.

Graham, F. D., *Exchange, Prices, and Production in Hyper-inflation Germany, 1920–23*, New York: Russell & Russell, 1930.

Jessen et al., Raymond J., "On a Population Sample for Greece," *Journal of the American Statistical Association*, September 1947, *42*, 357–84.

Okun, Arthur, "Efficient Disinflationary Policies," *American Economic Review Proceedings*, May 1978, *68*, 248–352.

Patterson, Gardiner, "The Financial Experiences of Greece from Liberation to Truman Doctrine (October 1944–March 1947)," unpublished doctoral dissertation, Harvard University, 1948.

______, personal interview, Washington, D.C., July 10, 1983.

Sargent, Thomas, "The Ends of Four Big Inflations," in R. Hall, ed., *Inflation: Causes and Effects*, Chicago: University of Chicago Press, 1982, 41–97.

Tobin, James, "Stabilization Policy Ten Years After," *Brookings Papers on Economic Activity*, 1: 1980, 19–72.

Bank of Greece, *Annual Report*, Athens, 1948.

High Board of Reconstruction, *National Income and Investment in Greece During the Years 1945–49*, Athens, March 1950, p. 2.

The Disinterest in Deregulation

By ROBERT E. MCCORMICK, WILLIAM F. SHUGHART II, AND ROBERT D. TOLLISON*

In the analysis of the costs of monopoly power, the usual experiment is to convert a competitive industry into a monopoly and observe the consequent change in consumer's surplus. Modern contributions have emphasized the deadweight cost of monopoly (Arnold Harberger, 1954) and the possibility of an associated rent-seeking cost of monopoly (Gordon Tullock, 1967). The Harberger cost, of course, refers to the lost consumer's surplus triangle; the Tullock cost concerns the role of competition for monopoly returns. Taken together and assuming that the competition for monopoly rents is perfect, the total cost of monopoly power is a trapezoid, the rectangle of monopoly profits plus the triangle of lost consumer's surplus (Richard Posner, 1975).

In this paper we approach the monopoly problem in a different spirit. We compare three states of the world—competition, regulation, and deregulation. In this setting we ask, what happens if a monopoly is eliminated through deregulation? Our analysis suggests that because under most conditions Tullock costs cannot be recouped, the returns to deregulation are lower than previously thought. Rent-seeking expenditures in the past leave the economy permanently poorer even if competition is restored to the industry. In contrast, the returns to preventing monopoly in the first place are relatively high in our model.

An insight afforded by the analysis is an explanation of the persistence of laws and regulations which appear to serve no interest. In this regard the example of railroad regulation in the United States comes to mind. The standard explanation for such regulation is either that voters and government decision makers are ignorant of basic economics or that a small interest group like railroad firms wins rents at the expense of uninformed or economically rational consumers of rail services who do not find it cost effective to seek deregulation. We offer another and perhaps more plausible explanation for the persistence of regulation and the apathy of consumers about the costs of regulation. Namely, the costs of such regulations are, for the most part, the original rent-seeking expenditures that lead to the regulation in the first place, and these costs are *sunk*. Abolishing so-called uneconomic laws does nothing to recover these losses. Hence, there is little political support from any quarter to return to the status quo ante. In fact, as we shall show, such a deregulatory program can easily impose more costs than it is worth.

There are numerous examples of this point, including tariffs and quotas of all sorts, subsidies to farmers, the postal monopoly, organized labor's antitrust exemption, the licensing of doctors, and so forth. The traditional explanations of these monopoly rights, namely ignorance of economic common sense and special-interest groups, are neither sufficient nor necessary. Since the primary costs of these laws are rent-seeking expenditures which are made prior to their passage, there is simply little to be gained by changing them now. Gains would accrue in the form of reduced Harberger costs; costs would be borne in passing and implementing the deregulatory program. It is not that the potential gainers from deregulation are large in number, diffuse, heterogeneous, and face high organizational costs, rather, *they do not exist to any degree*.

Interpreted in this light, efforts by political action groups, such as the Right-to-Work Foundation, stand to be a drain on society's resources. They cannot produce anything unless they prevent further monopolization through regulation. We do not, of course,

*McCormick and Shughart: Clemson University, Clemson, SC 29631; Tollison: Center for the Study of Public Choice, George Mason University, Fairfax, VA 22030. Thanks go to James Buchanan, Rex Cottle, and Gordon Tullock for helpful comments. The usual caveat applies.

expect these organizations to be ignorant of this principle. From a positive perspective, we expect them to allocate their efforts to prevent new regulations from appearing rather than to get old regulations abolished. Thus, groups such as the CATO Foundation and the Sierra Club will spend the bulk of their efforts on upcoming legislation, not on attempts to repeal old laws, even though those laws may have caused large losses to fiscal conservatives or to environmentalists.

I. The Economics of Deregulation

Figure 1 shows the monopoly situation we want to analyze.[1] The monopolist sells Q_m units per time period at price P_m. Total costs of monopoly and monopolization are the sum of the Harberger costs, H, and the Tullock rent-seeking costs, T.

The monopoly profits are competed away, perfectly here by assumption, through competitive attempts to capture them. That is, a priori the expected profits of being a monopolist through regulation are zero. And, the important point is that these expenditures are made before the monopoly is created. They do not simply vanish for a while; they are lost to the economy forever. Rent-seeking expenditures are therefore fundamentally different from H-costs. The T-costs cannot be recouped; they are removed from the production possibilities of the economy.

We have assumed for simplicity that the monopoly privilege is perfectly durable. Consequently, not only are the current rents competed away, but also the present value of all rents expected in the future. This need not be the case, of course.[2] If the monopolist must continue to compete with his rent-seeking rivals to maintain the status quo after he is regulated and protected from competition, then each period's T-costs will be smaller. In the nondurable case the monopolist must continually spend some resources to retain his protected status. Deregulation in this case

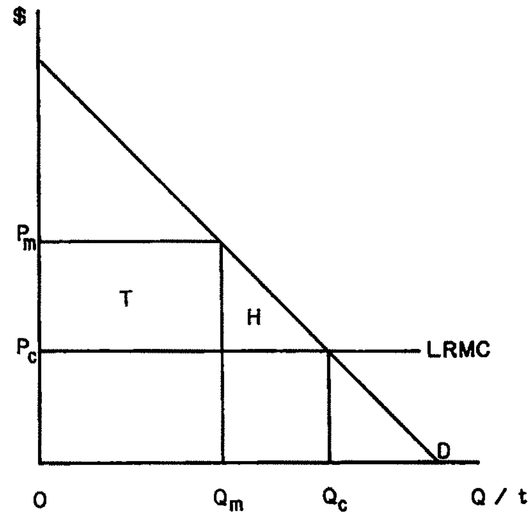


FIGURE 1

(provided it is durable) pays bigger dividends in terms of freed resources. We analyze the nondurable situation in more detail below. For the present we assume that all rent-seeking expenditures are made up front prior to the enactment of the monopoly regulation.

What are the implications of this analysis for converting an existing monopoly into a competitive industry? In terms of Figure 1 (some of) H can clearly be recaptured for the economy while none of T can. It is useful to think of what happens in terms of Figure 2 where the general equilibrium effects of monopolization and deregulation are shown.

The production possibilities frontier P_1 represents the economy in the initial state. With competition both in industry A and all other sectors of the economy B, the economy operates at a point such as 1, producing outputs Q_C^A and Q_C^B, respectively. If T-costs do not exist, the output restriction associated with monopolization of industry A moves the economy along P_1 to point 2, where output in the remaining competitive sectors of the economy has been augmented by the resources released from A. Although still on the P_1 frontier, point 2 is Pareto inferior to point 1 because of H-costs. In the standard analysis, deregulation could conceivably move the economy back to point 1, the status quo ante.

[1] We omit the marginal revenue function to avoid cluttering the diagram.

[2] See William Landes and Posner (1975) for an analysis of this point.

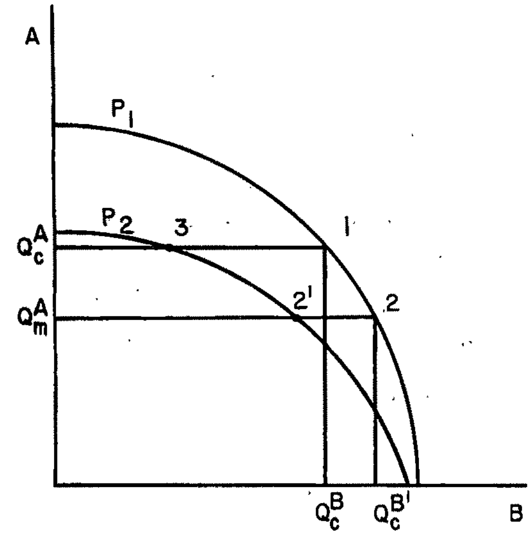


FIGURE 2

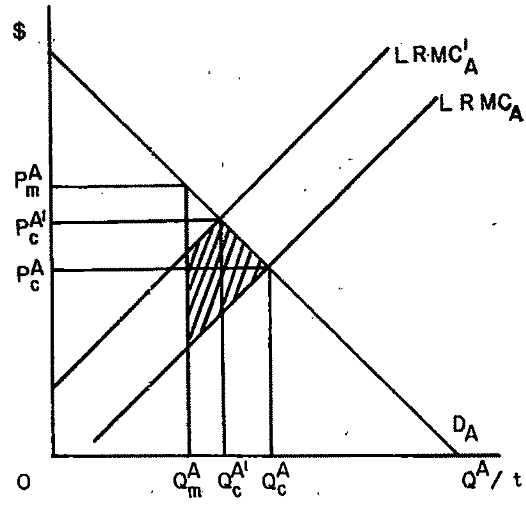


FIGURE 3

With rent seeking, however, monopolization of industry A moves the economy from point 1 to point 2′.[3] The production possibilities frontier shifts inward to P_2 as a function of T, the fixed costs of monopolization that are lost forever. Moreover, the inward shift is more pronounced in A the more the burden of the T-cost falls on that industry. Such an asymmetry would occur if the resources used in rent-seeking activities are specialized to A, or if for any reason regulation is nonneutral. Only by coincidence would P_2 be parallel to P_1. In any case, the status quo under regulation, point 2′, reflects the fact that as the result of T-costs, fewer resources are released by the output restriction in A to increase production in the other sectors of the economy.

Efforts to deregulate industry A can no longer restore the economy to point 1. A return to the previous competitive output level in A requires a movement along P_2 to point 3. The impossibility of returning to point 1 is a reflection of the real resources consumed in rent seeking. Moreover, because specialized resources in A were used for rent seeking, it is now more costly (in terms of B) to produce A.

These increased costs are illustrated in Figure 3. The P_2 frontier implies that the monopolized sector now has a higher long-run marginal cost schedule. In partial equilibrium terms, $LRMC_A$ shifts to $LRMC_A'$. Thus, part of the previous Harberger triangle (the shaded part) is also lost to the economy. *In other words, with perfect dissipation of monopoly rents, the gross gain to the economy from deregulation is less than the area of the original deadweight cost triangle.* The competitive position in A retrievable through deregulation is $(P_C^{A'}, Q_C^{A'})$, not (P_C^A, Q_C^A).[4]

[3]As discussed below, because regulation with rent seeking raises long-run marginal costs in sector A, the economy would actually move to a point on P_2 to the right of 2′. This changes nothing essential in our analysis, and we have chosen not to complicate Figure 2 by showing this effect.

[4]By symmetry, regulation lowers costs in the rest of the economy. The resulting higher output in B does not rescue the Pareto inferiority of regulation, however, since such a position was available when the economy chose 1 in Figure 2. Because in a two-good world A and B must be substitutes, deregulation in A lowers demand for B. Output and price in the rest of the economy fall, and, in turn, this will reduce the demand for A below that shown in Figure 3. Several rounds will be required to work through the income and substitution effects set up by deregulation in A. In terms of Figure 2, the adjustment toward a new equilibrium proceeds northwesterly along P_2 to arrive ultimately at some in-

The economic intuition of our analysis is simple. The original act of rent seeking that created the monopoly employed, for example, the best and brightest legal minds in the industry. Moreover, legal skills were acquired that serve no purpose other than rent seeking. This reduces the human capital stock of lawyers with knowledge of industry practices such as contract negotiation and enforcement. When the industry is deregulated, the cost of obtaining additional output rises because legal inputs are now relatively less efficient at providing services to the industry.[5]

This is just an example of the argument. The same point applies to other resources, especially managerial expertise and entrepreneurship, which are deflected from productive pursuits into the original act of rent seeking. It is, for example, fairly common to hear arguments that regulation drains managerial talents, dulling initiative and the incentive to innovate. On this account, marginal costs rise because the managerial talents acquired under regulation—dealing with regulatory bureaucracies, for example —are not instantly adaptable to a deregulated environment.[6]

We are aware that our analysis depends on perfect dissipation of monopoly profits and on the existence of a durable monopoly right. That is, we have assumed that the monopoly right was originally granted in perpetuity or, what is the same thing, that the winner does not have to defend his victory periodically. If, on the other hand, such expenditures recur in order to protect the monopoly right, it is possible to recoup a portion of the T-costs through deregulation. However, the previously expended resources can never be recovered, and our point holds. The question of durability is only a question of degree.[7]

Our analysis also abstracts from the costs of the public resources used in regulation and deregulation. Clearly, if deregulation is costly, the gain from increasing output in a previously monopolized sector will be even lower than our model suggests.

II. Conclusions and Implications

We have shown that the returns to deregulation are much lower than those depicted in the orthodox literature. Specifically, when the profits associated with a monopoly right are dissipated by rent-seeking activities, the economy is permanently poorer by the value of the real resources used in capturing the transfer. These resources are removed from the production possibilities of the economy, making a return to the status quo ante impossible. In addition, by drawing down the human capital stock of specialized inputs in the industry, rent seeking raises the monopolized sector's long-run marginal cost curve and engrosses part of the original Harberger triangle.

Harberger estimated the cost in the economy due to monopoly to be on the order of 1 percent of GNP. Taking into account Tullock costs, Posner suggested that monopoly costs were about 4 percent of GNP. Our basic point is that the majority of these costs cannot be recouped though efforts to get rid of monopoly and regulation. The gain from deregulation is less than the Harberger costs, perhaps one-half percent of GNP. An ounce of prevention is worth a pound of cure.

Finally, we return to the positive theme of the analysis. It should no longer be a mystery why there are, in fact, so few people who

termediate point between 2′ and 3. The result of deregulation will be an improvement in welfare, but the economy will be worse off than prior to regulation.

[5] More precisely, the stock of attorney human capital will have risen if the gains from rent seeking cause overinvestment in such skills. Again, however, this capital would have been used for rent-seeking purposes. When the industry is deregulated, costs will be higher than before because the attorneys find themselves forced into activities for which they have less basic training.

[6] The increase in marginal costs in the deregulated industry may be a short-term phenomenon. Over time, resources in the industry will adapt to the new competitive environment, and new resources coming into the industry will embody the requisite skills for working in a competitive rather than a government-protected sector. Thus, after the relevant adjustment period, the production possibilities curve in Figure 2 will rotate outwards to become parallel to the initial P_1 frontier, and $LRMC_A'$ in Figure 3 will shift to the right. Our basic point that you can't go home again in economic terms remains, however, because T-costs are permanently lost.

[7] We also note that the question of durability loses importance as the discount rate increases. At 15 percent the present value of $1 in ten years is only 25 cents.

stand to gain from regulatory reform. It does not take arguments about the lack of economic education or the impotency of consumers as an interest group to explain the persistence of economic regulation. There is simply little to be had from deregulation.

In fact, we offer something of a paradox of deregulation. The more an interest group has to spend annually to maintain its monopoly, the less durable it is by definition. More importantly, these expenditures signify that the present value in perpetuity of the T-costs has not been completely borne. By contrast, had the same monopoly been durable from the outset, the T-costs would have been expended prior to regulation, and the gains from deregulation would be less. Therefore, the more an interest group has to spend to maintain its monopoly (holding constant the monopoly rents), the greater the gain from deregulation and the more likely deregulation will occur.

REFERENCES

Harberger, Arnold, "Monopoly and Resource Allocation," *American Economic Review Proceedings*, May 1954, *44*, 77–87.

Landes, William M., and Posner, Richard A., "The Independent Judiciary in an Interest-Group Perspective," *Journal of Law and Economics*, December 1975, *18*, 875–901.

Posner, Richard A., "The Social Costs of Monopoly and Regulation," *Journal of Political Economy*, August 1975, *83*, 807–27.

Tullock, Gordon, "The Welfare Costs of Tariffs, Monopolies, and Theft," *Western Economic Journal*, June 1967, *5*, 224–32.

Market Opportunities, Intrafamily Resource Allocation, and Sex-Specific Survival Rates: An Intercountry Extension

By RATI RAM*

In a recent work, Mark Rosenzweig and T. Paul Schultz (1982) developed a model of intrafamily resource allocation, and provided evidence from Indian data suggesting that differences in male-female child survival rates are partly attributable to the economic opportunity differentials for the two sexes; more specifically, a weaker economic opportunity set for adult females induces a smaller allocation of family resources toward female children and thus (along with other factors) leads to a lower survival rate for female children than for males.

The objective of this paper is to extend the evidence in two directions. First, instead of looking at only sex-specific *child* survival rates, one may consider the overall survival rates for males and females since the structure of intrafamily resource allocation for adults should also follow the same basic logic as that toward children. Second, it seems useful to conduct the study for countries covering a wider range than the Indian subcontinent. In the Indian case, an intuitive basis for the model is easily suggested by the observation that females have lower participation rates and also lower survival rates than males. However, in almost *all* other countries of the world, a solid, but perhaps not widely recognized, fact is that females have typically longer lifespans but weaker economic opportunities as measured by the market participation rates. It is probably of greater interest to examine whether in the latter contexts economic opportunities exercise a (partial) influence on resource allocation to males and females in the direction suggested by basic economic theory.

*Illinois State University, Normal, IL 61761. Insightful comments by an anonymous referee of this *Review* are gratefully acknowledged. Rosa Lea Danielson provided useful research assistance. Responsibility for any errors or deficiencies is mine.

This study uses three data sets. One is a cross section of 118 countries for 1970; the second consists of a cross section of states in the United States around 1970; and the third contains 32 annual observations for the United States covering the period 1947–78. The main conclusion is that market opportunity differentials, as proxied by participation rates, do seem to affect the allocation of resources, as measured by survival rates, in the direction suggested by economic theory; in particular, an increase in the market participation rates of females enhances the survival differential in their favor.

I. Elements of a Theory

The basic theoretical model underlying this study is the same as that of Rosenzweig and Schultz. In the context of intrafamily resource allocation, it is postulated that *one* of the determinants of the allocation of family resources toward any member is the character of the present and future market opportunities for the member; *ceteris paribus*, stronger market opportunities induce greater allocation of family resources toward the member. Such a postulate seems plausible because stronger market opportunities are likely to imply a higher return on the resources, and the family as a unit may want to follow the postulated pattern either because of interdependent utilities or because of members' earnings being shared or both. In this study, as in Rosenzweig-Schultz, the focus is on the allocation of family resources toward male and female members. However, while their work analyzed the resource-allocation differentials (as measured by sex-specific child survival rates) with respect to male and female children, this study investigates resource-allocation differentials relative to all male and female members. Therefore, instead of relating sex-specific child survival patterns to male-female participation rates

(market opportunities), the relation between sex-specific market participation rates and male-female differentials in *lifespans* is explored. The argument that sex-specific market participation rates affect intrafamily resource allocation toward male and female adults is at least as persuasive as that such participation rates affect family resource allocation patterns relative to male and female children.[1]

Since intrafamily resource allocation patterns are not easily observed in a direct manner, the empirical proxy for these is the relative survival pattern for the two sexes, which is assumed to be determined (among other things) by intrafamily resource allocation. Thus, the estimating equation employed by Rosenzweig-Schultz (p. 813) for the district level analysis is adopted here with minor modifications, and the model may be simply written as

$$DS_{fm} = g(E_f, E_m, Y), \tag{1}$$

where DS_{fm} represents female-male survival differential, E_f is a variable measuring the "economic opportunities" for females, E_m represents similar opportunities for men, and Y is an income term.

Although by no means complete or perfect, the specification seems workable for the main purpose of the study, namely to judge if economic opportunities do exercise an influence on allocation of family resources to the two sexes in the manner suggested by theory. The expected signs on the partial derivatives with respect to E_f and E_m are a straightforward matter; E_f should have a positive effect on female-male survival differential, and E_m should have a negative effect at least in terms of the simplest behavioral logic. The sign on the coefficient of the income term is ambiguous. Although an increase in income is expected to increase the allocation of resources to, and thus the survival rates of, both sexes, the effect of higher income on survival *differential* (which is what DS_{fm} denotes) is not obvious.

Linear and double-log versions of (1), with appropriate empirical proxies for DS_{fm}, E_f, E_m, and Y, are employed in the empirical work.

II. Intercountry Data and the Main Results

The empirical proxies for E_f and E_m are the market participation rates of females and males respectively. Real (PPP) GDP per capita of Irving Kravis et al. (1978) is used for Y. Female-male differential in life expectancy at birth is used as a proxy for DS_{fm}. Life expectancy at birth (lifespan) evidently reflects well the aggregate survival rates for each sex, and thus seems to be a good proxy. In the linear model, the excess of female lifespan over the male lifespan is used as the dependent variable, and a Y-square term is added to capture any nonlinearity in the effect of Y. In the log-log model, difference of the logarithms of female and male lifespans (or the logarithm of their ratio) is the dependent variable. In regressions that include both developing and industrialized countries, a DC "dummy" has been added.[2]

One might ask whether the participation rate of all females is an appropriate empirical proxy, and whether participation rate of married females might not be a better variable.[3] It seems that participation rates of all females (and all males) would be more appropriate than participation rates of married females (and married males). When the focus is on sex-specific allocation of family resources toward children, the appropriate signal about market opportunities is obviously provided by the participation rates of *all* females (and all males). Even when one considers the allocation of family resources toward adults, including oneself, the participation rates of all females (and all males) would appear to provide the relevant signals regarding market opportunities and hence regard-

[1] One advantage of working with overall sex-specific survival rates, rather than with child survival rates, is that data availability is much better for most intercountry samples.

[2] See the notes in Table 1 for more details concerning the variables.

[3] This point, and the one about the effect of female participation rate operating through maternal mortality, were suggested by an anonymous referee of this *Review*.

ing allocation of family resources. For example, considering the case of an adult female, whether married or single, the participation rate of all females, and not merely of married females, should signal her likely market opportunities, and hence the optimal level of family resource allocation toward her. Therefore, it seems appropriate to relate variables reflecting family resource allocation with overall sex-specific participation rates which reflect market opportunities for each sex.[4]

The intercountry sample consists of 118 countries, of which 17 are "industrialized" and the other 101 "developing." The variable values are centered on the year 1970.[5]

Table 1, part I, contains the main intercountry results based on estimation of linear and logarithmic versions of (1). It is observed that in each of the six regressions, the estimated coefficient of the female participation rate variable has the expected positive sign, and in four of the six cases the estimate is significant at the 1 percent level or better. The low significance of the estimates for industrialized countries might be due to the very small size of the subsample. The evidence thus supports the view that higher female participation rates, that presumably signal better market opportunities for females, tend to shift allocation of resources toward them and that shift is reflected in longer female lifespans.

The coefficients of the male participation rate are somewhat mixed. Of the six estimates, four have the expected negative sign, but only two are significantly negative. These results are similar to those reported in the second row of Table 6 in Rosenzweig-Schultz (p. 813). Several explanations can be suggested. First, male participation rates are uniformly high in most countries. The coefficient of variation for male participation rates in the sample is 0.1 as opposed to 0.6 for the female rate. Thus although most regression estimates display the right sign, the estimates may not be sharp enough due to a low sample variance of the regressor. Second, while a higher participation rate by men may indeed shift family resource allocation toward them, an increase in their work efforts from the already high levels, *and* the kind of market work they do, might increase what is loosely called "job stress" and may offset some of the effect of increased resource allocation; thus the *observed* survival rate may move only slightly in their favor.

The effect of income is observed to be nonlinear. An increase in income appears to increase the female survival advantage but at a declining rate.

The coefficient of multiple correlation in several regressions is fairly good; but in some it is rather low. It is certainly possible that some "relevant" variables are not included; indeed the model does not claim to be an explanation of sex-differentials in survival rates; it seeks to isolate mainly the effect of economic opportunities on family resource allocation as reflected in survival rates. Thus, due to the simple structure of the model, one might feel uneasy about the estimation procedure (the method of ordinary least-squares), even though one would hardly expect a major feedback from life expectancy differential to participation rates. Also, one might suspect some heteroscedasticity in the error term and thus loss of efficiency.

A test recently proposed by Halbert White (1980) seems to be a reasonable procedure for testing the joint hypothesis of homoscedasticity and lack of specification error (in the sense of regressors being independent of the disturbance term).[6] The simpler version of the test (White, p. 824) essentially consists of regressing squared *OLS* residuals on squares and cross products of the original regressors, and testing nR^2 as $\chi^2_{K(K+1)/2}$ where n is the number of sample observations, R^2 is the coefficient of determination in the test regression, and K is the number

[4]Also, as the discussion in Section IV indicates, participation rates of all females and married females seem to yield estimates that have similar patterns.

[5]A list of the sample countries is available on request. The terminology used in World Bank (1981) is adopted for classifying countries into developing and industrialized groups. Every country for which data are available in the cited sources has been included in the analysis.

[6]Of course, other tests are also available for investigating specification errors of this kind. However, White's test appears simple and good in the sense it does not depend on instrumental variables, the choice of which can seriously affect the results.

TABLE 1—PARAMETER ESTIMATES IN REGRESSIONS OF FEMALE-MALE SURVIVAL (LIFESPAN) DIFFERENTIALS

	FPARTICIP	*MPARTICIP*	*Y*	*YSQ*	*CD*	*N*	$\hat{\rho}$	R^2
I. Intercountry Sample								
A. Linear Model[a]								
i. All Countries	.0261	−.0168	.0949	−.0007	.9473	118		.46
	(2.52)	(−.60)	(4.61)	(−2.89)	(1.40)			
ii. Developing Countries	.0252	−.0003	.1178	−.0011		101		.23
	(2.29)	(−.01)	(4.28)	(−2.73)				
iii. Industrialized Countries	.0361	−.1410	.0827	−.0005		17		.34
	(1.01)	(−2.30)	(.65)	(−.60)				
B. Logarithmic (log-log) Model[a]								
i. All Countries	.0083	.0187	.0060		.0067	118		.16
	(2.64)	(.67)	(1.61)		(.64)			
ii. Developing Countries	.0082	.0270	.0063			101		.08
	(2.42)	(.89)	(1.57)					
iii. Industrialized Countries	.0197	−.1211	.0041			17		.33
	(1.27)	(−2.40)	(.27)					
II. U.S. Data								
A. Linear Model[a]								
i. Cross Section (about (1970)	.0809	−.0904	−.00004	−.0005		51		.36
	(3.04)	(−2.67)	(−.03)	(−.26)				
ii. Time-Series[b] (1947–78 annual)	.0438	−.0225	.0048	−.0062		32	.16	.99
	(2.34)	(−1.09)	(11.51)	(−12.02)				
B. Logarithmic (log-log) Model[a]								
i. Cross Section	.0623	−.1294	−.0248			51		.35
	(3.32)	(−2.97)	(−2.16)					
ii. Time-Series[b]	.0023	−.1593	.0136			32	−.69	.83
	(.10)	(−2.86)	(.83)					

Notes: t-statistics are shown in parentheses. 1) The variable definitions for the intercountry sample are: *FPARTICIP* and *MPARTICIP* are females and males in the labor force as percentages of their respective populations; *Y* is the real (*PPP*) *GDP* per capita index of Kravis et al. (with U.S. = 100); *YSQ* is the square of *Y*; *DC* takes the value 0 for developing countries and 1 for industrialized countries. All variables are for a period around 1970; life expectancies are for 1970–75. Data sources are: life expectancies, United Nations (1981); participation rates for developing countries, World Bank (1981); *Y*, Kravis et al.; participation rates for industrialized countries, OECD. 2) For the U.S. time-series data, *Y* is disposable personal income per capita in 1972 dollars, and participation rates are of males and females aged 16 and over. For the U.S. cross-section data, *Y* is per capita personal income for 1970, and life expectancies are for 1969–71. Data sources are fairly standard, and details can be obtained from the author. District of Columbia is one of the observations in the U.S. cross section. 3) The constant term *is* included in all regressions, but its estimates are not reported to save space.

[a] The dependent variable in the linear version is the excess of female life expectancy over the male life expectancy (at birth in years). In the logarithmic versions, the dependent variable is the logarithm of the ratio of female and male life expectancies, and all variables on the right hand side (except *DC*) are logarithmic transformations.

[b] The reported numbers are the maximum likelihood estimates premised on a first-order autoregressive error term (estimation in all other equations being by the ordinary least squares method).

of original regressors. White's test was conducted on the specifications for which estimates are reported in lines I.A.i and I.B.i of Table 1 (without the *DC* dummy). In neither case the hypothesis of homoscedasticity and lack of specification error could be rejected at the 5 percent level.[7] Thus, one can be a little more confident about the message of the intercountry results in Table 1 since there seems no evidence of a major specificational problem.

III. Nonparametric Evidence from Intercountry Data

Although Table 1 indicates reasonably well the intercountry situation for a recent cross section, one may also wish to observe the change in the position over the last two or three decades during which female participa-

[7] The values of χ^2 are 9.1 and 11.2 for the two equations, while the critical values (at 5 percent) are 23.7 and 18.3, respectively. Details are available on request.

TABLE 2—MALE (M) AND FEMALE (F) LIFE EXPECTANCY IN VARIOUS PARTS OF THE WORLD (in years; 1950–55 and 1970–75)

	1950–55		Excess of F over M	1970–75		Excess of F over M
	F	M		F	M	
World	48.4	46.0	2.4	57.1	54.6	2.5
Developed Countries	68.7	63.0	5.7	74.7	67.8	6.9
Less Developed Countries	43.2	41.6	1.6	54.0	52.3	1.7
Africa	38.7	35.9	2.8	47.9	44.9	3.0
Latin America	52.7	49.7	3.0	62.7	58.8	3.9
North America	72.0	66.3	5.7	75.3	67.7	7.6
East Asia	49.0	46.0	3.0	65.3	62.3	3.0
East South Asia	40.6	38.2	2.4	50.9	47.7	3.2
Mid South Asia	38.3	39.5	−1.2	46.5	47.6	−1.1
West South Asia	45.8	43.7	2.1	56.6	54.6	2.0
Europe	67.7	63.2	4.5	74.1	68.4	5.7
USSR	68.5	60.0	8.5	74.3	65.5	8.8

Source: United Nations, pp. 82–100.

tion rates have risen rapidly in many countries. Although time-series data on variables of interest seem rare, and thus parameter estimates from such data are hard to obtain, the simple, nonparametric evidence presented in Table 2 appears quite suggestive. Two things are clearly seen from the data given in the table: first, in almost every part of the world, females have had an advantage over males in lifespan; second, in practically every case, the female advantage has increased between 1950–55 and 1970–75, which is also the period of a substantial rise in female participation rates in most countries. In particular, the increase in the female lead in lifespan has been the largest in North America, where the rise in the female participation rate over the period has probably been the most dramatic.

The fact of females having had a lead in lifespan over males in most of the world, and the observation of their lead having increased over the last three decades, also suggests an interesting implication which is not directly related to the main subject of this study. If, as appears to be widely accepted, lifespan is a major indicator of well-being or "quality of life," females have been better off than males in most of the world, and, during the last 25 to 30 years, their welfare level has improved relative to males in almost every country of the world. The implication appears contrary to what is commonly believed about the level of female well-being relative to that of males.

IV. Evidence from the United States

Since female labor force participation in the United States has increased quite rapidly during the last about 30 years, and as Table 2 shows a substantial widening of the female lead in lifespan in North America, it is tempting to study the U.S. case a little more carefully. Table 1, part II, contains estimates of the linear and log-log variants of equation (1) with respect to two data sets for the United States, one of which is a cross section of states around 1970, and the other consisting of 32 annual observations from 1947 through 1978. The *OLS* estimates are reported for the cross section, and maximum-likelihood estimates with a first-order autoregressive disturbance term are given for the time-series data set. It is obvious that the evidence from the U.S. data also supports the hypothesis that an increase in the female market participation rates tends to shift the family resource allocation toward them and thus enhances their lead in lifespan over the males.[8]

[8]To save space, estimates obtained from the U.S. data are discussed very briefly. The pattern of the coefficients of the income terms is understood easily once it is

An important question, applicable to both the U.S. and the intercountry estimates, is whether there might be some factor other than intrafamily resource allocation that could explain the observed relationship between the female participation rate and the female-male life expectancy (mortality) differential. Although there are several minor linkages that can be thought of, there is one that needs explicit mention. It is likely that increased female participation in market work reduces lifetime fertility and thus maternal mortality. While it might seem that maternal mortality is perhaps only a minor component of the total female mortality, and thus exercises a small influence on female lifespans, the aspect merits some empirical exploration. Since reduction of maternal mortality caused by higher labor force participation would occur primarily among married females in most countries, one approach can be to see whether market participation rates of *married* females do better than the overall female participation rates. However, since participation rates by martial status are not easily available for many *LDC*s, the approach is difficult to implement for the intercountry sample, but can be tried for the United States. Regressions of U.S. male-female age-adjusted mortality rate differentials over 1955–80 in an equation like (1) indicate that the overall female participation rate does almost as well as the married female participation rate in terms of R^2 and standard error of the regression (SEE).[9] Therefore, it does not seem likely that changes in maternal mortality affect the estimated coefficients of the participation rate variables in any significant manner. In fact, a more direct test of the issue can be conducted by introducing a fertility rate variable in equation (1). However, even when that is done, the pattern of the estimated coefficients of the participation rate variables remains unaltered.[10]

V. Concluding Remarks

This study extends the recent work of Rosenzweig-Schultz by taking a larger intercountry sample and by considering the overall female-male survival rates as related to economic opportunity differentials for the two sexes. Despite the mixed quality of the data and the simplicity of the model and estimating procedure, the evidence seems to support fairly well the proposition that the partial effect of better market opportunities for females, as reflected in higher participation rates, is to shift allocation of family resources toward them and thus to generate higher female survival rates and longer female lifespans. One might ask whether these results are contrary to the beliefs concerning the influence of "altruism" in intrafamily resource allocation (for example, Gary Becker, 1981). Perhaps not. Probably it is not "selfishness" but economic rationality that induces allocation of more resources in the direction that provides a higher economic payoff. It is perhaps not the case that this evidence, or that provided by Rosenzweig-Schultz, would necessarily tend to alter one's belief about the extent to which love and affection guide family resource utilization; it only seems to say that love and affection are

noticed that the cross section of 1970 is based on a relatively high income level at which increased income no longer enhances the female advantage in lifespan; but the time-series data set covers a wide income range over which higher income does (as in the intercountry sample) tend to increase the female lead in lifespan.

[9]In the linear version, the R^2 and SEE are the same for overall female participation rate and married female participation rate. In the logarithmic version, R^2 and SEE are .97 and .0097 with overall female participation rate, and .96 and .0101 with married female participation rate.

[10]For example, the intercountry logarithmic regression like Table 1, (I.B.i), yields the following estimates when GRR (gross reproduction rate) is added (t-statistics are in parentheses):

$$\begin{aligned} LDIFF = &\underset{(-.30)}{-.037} + \underset{(2.42)}{.0081} LFPARTICIP \\ &+ \underset{(.56)}{.0163} LMPARTICIP + \underset{(1.06)}{.0052} LY \\ &+ \underset{(.51)}{.0057} DC \; \underset{(-.27)}{-.0030} LGRR. \end{aligned}$$

It is obvious that above estimates have the same pattern as those in line I.B.i, Table 1.

not inconsistent with economic rationality. As an incidental aspect, Table 2 shows that, if life expectancy at birth is regarded as an important indicator of well-being, females have had a higher level of welfare than males in practically every country, and their advantage has increased in all parts of the world during the last three decades.

REFERENCES

Becker, Gary S., "Altruism in the Family and Selfishness in the Market Place," *Economica*, February 1981, *48*, 1–15.

Kravis, Irving B., Heston, Alan W. and Summers, Robert, "Real *GDP* Per Capita for More Than One Hundred Countries," *Economic Journal*, June 1978, *88*, 215–42.

Rosenzweig, Mark R. and Schultz, T. Paul, "Market Opportunities, Genetic Endowments, and Intrafamily Resource Distribution: Child Survival in Rural India," *American Economic Review*, September 1982, *72*, 803–15.

White, Halbert, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, May 1980, *48*, 817–38.

Organisation for Economic Cooperation and Development, (OECD) *Labour Force Statistics 1968–1979*, Paris, 1981.

United Nations, *World Population Prospects as Assessed in 1980*, New York, 1981.

World Bank, "Social Indicators Data Sheet," mimeo., Washington, May 1981.

______, *World Tables 1980*, Baltimore: Johns Hopkins University Press, 1980.

Directly Unproductive Prophet-Seeking Activities

By AVINASH DIXIT AND GENE GROSSMAN*

In this note we merge two themes that have been prominent in recent years in economics. The first provides economic explanations of seemingly noneconomic phenomena. The other studies the allocation of resources to the activities of creating and capturing rents in the economy. Thus we have the economics of sociological matters on the one hand, and the politics of economic allocation on the other. We study a problem that involves both transformations, namely the political economy of organized religion.

An initial exploration of such a deep subject must make drastic simplifications to achieve any progress. We do this by constructing a pilot model that resembles familiar ones as closely as possible.

Consider an economy whose "worldly" component produces two goods, labelled guns and butter for convenience, using two factors, capital and labor. With given endowments of factors, and for the moment omitting any "other-worldly" uses of their services, we have a conventional transformation curve *TT*. Let p^* be the world price of guns in terms of butter, again assumed constant for expository convenience. Then production will occur at the point *A*. This is shown in Figure 1.

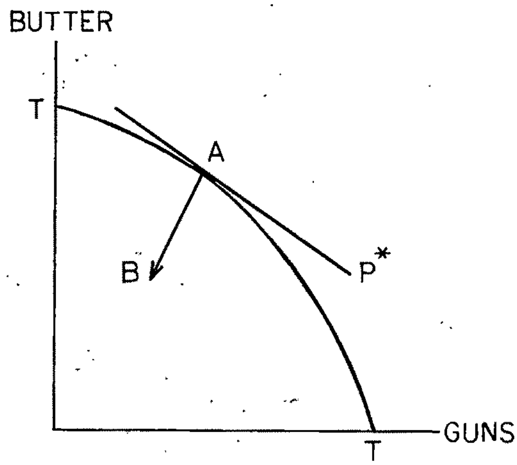


FIGURE 1

Now we introduce the next world, but for the moment leave out organized religion. We do this by introducing a heaven, which consists of a fixed number $\mathring{N}$ of places. (Halos over a symbol will identify heavenly magnitudes.) These are available to the population of the economy in the afterlife. We assume that $\mathring{N}$ falls short of the population, therefore there will be a scarcity rent or shadow price of each space. We assume for simplicity that each member of the population would be willing to part with $\mathring{p}$ units of present-value butter to secure a place in heaven. This reservation price will then equal the rent; to distinguish it from the world price, we refer to it as the other-world price.

Suppose first that entitlements to places in heaven are awarded to $\mathring{N}$ members of the population by an exogenous force. This could be the deus ex machina, or the invisible hand. The important point here is that the mechanism cannot be influenced by the population, and no resources need be spent in seeking activities. The production will remain at *A*, and there will be some lucky recipients of tickets to heaven. This is the first best, the situation described by Dr. Pangloss in *Candide*, and we shall label it Nirvana.

Next suppose it becomes possible to win places in heaven by spending resources in this world. We interpret this in the obvious way as organized religion, namely an intermediary of this world who acquires the rights to distribute tickets to heaven. We make the usual rent-transformation assumption that the religious sector grows until payments to factors there absorb the available rents. Implicitly, we require that there be free entry

*Princeton University, Woodrow Wilson School of Public and International Affairs, Princeton, NJ 08544. Financial support for this research was provided by pennies from heaven.

into the religion business, an assumption that seems well supported by casual empirical observation.

Formally, the world price p^* fixes the factor prices w, r through the usual unit-cost relations for guns and butter. This in turn fixes the input coefficients a_{ij} for $i = L, K$ the inputs and $j = G$, B, and R the outputs. The equilibrium conditions to determine the output levels are

$$a_{LG}G + a_{LB}B + a_{LR}R = L;$$

$$a_{KG}G + a_{KB}B + a_{KR}R = K;$$

$$(wa_{LR} + ra_{KR})R = \mathring{p}\mathring{N}.$$

Geometrically, the withdrawal of resources $(a_{LR}R, a_{KR}R)$ from worldly production activities moves the production point from A to B along a "Rybczynski" line, as shown in Figure 1. Welfare has clearly fallen, since the value of worldly goods has decreased and that of places in heaven is unchanged.

The above analysis assumes this world to be perfect. Alas, it is not. Several distortions can cause the production equilibrium to depart from the point of tangency with p^*. Guns might be underproduced if they are a public good with demand revelation problems. They might be overproduced if there are Prisoner's Dilemma problems among different nations of this world. When the world is replete with such economic sins, the diversion of resources into religious activities may be welfare-improving in the standard manner of second-best paradoxes.

Such a possibility is illustrated in Figure 2. Suppose the equilibrium without organized religion is at C, with overproduction of guns. Suppose guns are very capital intensive relative to butter. Now introduce an organized religion whose technology is also very capital intensive, involving large temples or cathedrals and ornate and expensive artifacts. Movement of resources into it will shift the production of worldly goods along an upward-sloping Rybczynski line CD. It is perfectly possible for this to be steeper than the world price line p^* through C, as shown. Then welfare will be higher with the religion than without. Conversely, if butter (the

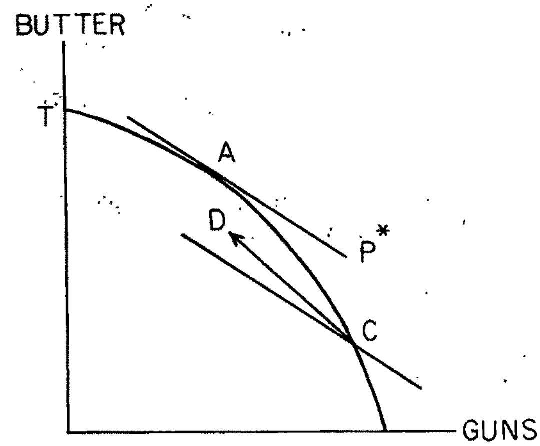


FIGURE 2

labor-intensive good) were being overproduced, it would be conceivable that a very labor-intensive religion relying on meditation, etc., might provide a second-best welfare improvement. Of course, the model rules out achievement of Nirvana even via meditation.

We recognize that the model as it stands is too simple to allow us to draw dogmatic conclusions. The following extensions suggest themselves for future research.

1) The willingness to pay for a place in heaven can be endogenized by the introduction of promotional activities. This will allow us to model rent creation as well as rent-seeking religious activities.

2) The recognition that different religions have different attributes naturally suggests a model based on product differentiation and monopolistic competition.

3) A hell could be introduced. This will allow us to model Edgeworth's concept of "economic damnification" in a more meaningful way. Students of the literature will be aware that Oscar Wilde suggested a similar three-location model (the choice between this world, the next world, and Australia), but did not conduct any formal analysis.

4) The result that religious activities can raise or lower welfare depending on the circumstances has obvious policy implications, and leads to the general question of separation of church and state. But this is too difficult an issue to be pursued here.

The Client Relationship and a "Just" Price

By A. GLAZER*

A consumer can be exploited by a firm's monopoly power even in markets consisting of many firms. This is likely to occur if the consumer is ignorant of the good or service he actually requires (so that he cannot search among firms prior to going to any one), and if high search costs force the consumer to purchase the good at the first firm he visits. These difficulties arise, for example, in the markets for automobile repairs, medical care, and legal advice.

The basic problem caused by the presence of search costs was first described by Peter Diamond (1971), who shows that the market structure will be monopolistically competitive rather than perfectly competitive. Diamond supposes that all consumers have the same reservation price, r, for a good, and that the cost of moving from one shop to another is s dollars. The only equilibrium price in such a market can be r dollars. For if firms charged a price r_1 which is lower than r, any one firm could charge a price of $r_1 + s$ and that firm's customers would have no incentive to leave that shop in search of another one. This argument can be applied to *any* price less than r, so that r must be the only equilibrium price.

An obvious limitation on a firm's incentive to charge a high price, even when each arriving customer purchases the good or service regardless of cost, is that the higher the price it charges, the fewer customers it will have. This can occur for at least two reasons. First, the firm might find that it can increase its price only by increasing the length of time it serves each customer. This is especially common in markets, such as those for taxicab, legal, or plumbing services, where payment is based on an hourly rate (see my paper with Refael Hassin, 1983).

Second, a customer may be reluctant to return to a firm that had charged him a high price in the past. This phenomenon has been the subject of a fair amount of study, and is the subject of the present paper. Michael Rothschild (1973), for example, studies a model in which price reputations play an important role: a firm that raises its price will lose some, though by no means all, of its customers. He finds that in such a market there exists a Nash equilibrium in which the price is less than the monopolistic price, but greater than the price that would prevail under perfect competition. Similar results are obtained by Edmund Phelps and Sidney Winter (1972).

Michael Darby and Edi Karni (1973) give a thorough discussion of the importance of repeat business. They discuss the market for automobile repairs in the light of "credence" qualities, defined as those qualities that are worthwhile but cannot be evaluated in normal use. In the case of repairs, credence qualities arise when consumers cannot easily judge whether a particular repair was actually necessary for the proper functioning of the broken item.

A firm can exploit such consumer ignorance by prescribing and performing unnecessary repairs. The profitability of doing so, however, is limited by the firm's concern that a customer may decide not to purchase any repairs at all, or that he might avoid returning to a shop which he feels had overcharged him. Darby and Karni's analysis implies that a firm is more likely to defraud tourists and casual customers than to defraud customers it believes might make purchases in the future. Similarly, it is less likely to swindle a customer who is quite certain of which repair he requires.

Darby and Karni elucidate the major issues, though unavoidably they could not cover them all in one paper. One neglected

*School of Social Sciences, University of California, Irvine, CA 92717. This paper is based on my dissertation submitted at Yale University. I am indebted to William Brainard, Richard Nelson, and Sidney Winter for many helpful comments. Financial support was provided by a grant from the Alfred P. Sloan Foundation.

issue will be of special concern here. In many markets, a consumer never knows whether he was overcharged for a repair; the consumer has no objective measure by which to judge the quality of the repair or the amount of work actually required to repair his machine. (In this respect, the present analysis differs from that given by Dennis Smallwood and John Conlisk, 1979). Instead, a consumer searches for another shop in the following period whenever he is charged a price greater than some critical one. This price can be termed as "just" or "fair." What is the equilibrium value of this critical price? Is a consumer's conception of what constitutes a fair price simply a creation of his imagination, or can it be based on some rational, utility-maximizing, calculus? What is the value of the just price that would maximize consumer welfare? Is this optimal just price identical to the market equilibrium just price? Should government try to influence consumers beliefs about what is a fair price?

These questions are studied below by means of a relatively simple model. Obviously, any such stylized model cannot capture all the characteristics of a market in which the client relationship plays a role. But it does capture and illuminate the essential features of such a market. For concreteness, the discussion is couched in terms of repairs, though the model can be interpreted to refer to other markets as well.

The assumptions used throughout the paper are spelled out in Section I. Section II presents the model and describes the equilibrium solution for a given value of the just price. Section III discusses the equilibrium value of the just price.

I. Assumptions

Let us assume that the high cost of moving a broken good from one shop to another deters a consumer from undertaking search. The consumer purchases the repair at the first and only shop he visits each time he needs a repair. The model can be easily generalized, with no change in the substantive conclusions, to allow for a downward-sloping demand curve: in that case a high price many cause some customers to refuse the repair in the current period as well as to search for a different shop in future periods.

Because of the need for a diagnosis, price estimates are obtainable only *after* the consumer has arrived at a shop. Neither before nor after he purchases the repair does the customer know what was wrong or how much it cost the firm to fix the item. A consumer can costlessly switch shops each time he purchases a repair: he need not buy them all at the same shop. Each consumer owns a good which breaks once in every period.

The next time he requires the services of a repair shop, the consumer can choose to return to the shop he visited last time, or he can go to a different one. Most generally, the probability that a consumer will return to some shop will be a function, $F(p)$, of the price he was charged there last time. We would expect $F(p)$ to be a decreasing function of p. Suppose the firm desires that a fraction, R, of its customers return next time.[1] If $F(p)$ is a concave function of p, the firm should charge all customers a price of $F^{-1}(R)$. If, however, $F(p)$ is a convex function in the vicinity of $F^{-1}(R)$, the firm should charge two different prices, p_1 and p_2 (where $p_2 > p_1$). The firm can charge a fraction, t, of its customers a price of p_2, and a fraction, $(1-t)$, a price of p_1.

As the latter is the more interesting solution, I shall concentrate on it. I make the following simplifying assumptions about this function $F(p)$. A customer returns to a shop in the next period if and only if the shop charged him p_1 or less for the repair in the previous period. The customer will not return to a shop which had charged him p_2; he will instead go to some shop chosen at random from all repair shops in the market. For the present, p_1 will be considered as exogenously fixed; in Section III it is treated as an endogenous variable.

The value of p_2, the higher of the two prices, can be viewed in several ways. It can be the consumers' reservation price. Alternatively, p_2 may be set by governmental

[1]The following section describes how such a value is determined.

decree, or perhaps by a manufacturer who regulates his dealers. Finally, p_2 can be endogenously determined as part of the firm's profit-maximizing calculus. For my purposes, however, the precise value of p_2 is unimportant, and I shall not analyze it further.

The present model assumes that the firm cannot distinguish between customers. It could well be, however, that a customer owning a new automobile is willing to pay more for a repair than is a customer owning an old one. This phenomenon can be incorporated into the model by considering each firm as serving several different markets—for example, a market for repairs of new cars and a market for repairs of old ones. This would mean that any firm charges more than two prices. The essential features, however, of the model would not be altered by explicitly considering this possibility, and it is therefore ignored hereinafter.

I use the following notation. Let q_i be the number of customers at a shop in period i, and M_i the number of new customers at a given shop in period i. These new customers are consumers who purchased a repair at some other shop in the previous period, were charged p_2 for the repair, decided to search for a new shop, and chose the given shop. The value of M_i is easily determined. Let Q be the total number of consumers in the market, N the number of shops, and f_{i-1} the fraction of all customers in the market charged p_2 in period $i-1$. In period $i-1$, Qf_{i-1} customers were charged p_2 and these customers search for a new shop in the following period. Of these Qf_{i-1} customers, $Qf_{i-1}/(N-1)=M_i$ arrive at any specified shop in period i.

Suppose there are a sufficiently large number of consumers and shops in the market so that M_i is not affected by a firm's own behavior. For simplicity suppose also that the effects of random variations in the number of customers at each shop are insignificant, and can be ignored. All firms are identical, with free entry and exit in the market.

II. The Firm's Policy

A firm's objective is to maximize the discounted value of its profit stream, or to maximize

$$(1)\quad \sum_{i=1}^{\infty}(1+r)^{-i}\{[p_1 + t_i(p_2-p_1)]q_i - C(q_i)\},$$

subject to the condition that

$$(2)\quad q_{i+1}=(1-t_i)q_i+M_{i+1}, \qquad i=1,2\ldots,$$

where M_{i+1} is the number of new customers who arrive at the firm in period $i+1$, t_i is the fraction of customers that the firm charged a price of p_2 in period i, $(1-t_i)$ is the fraction of customers it charged p_1, r is the firm's discount factor, and $C(\cdot)$ is the firm's cost function. The firm treats r and M_i as fixed; its choice variables are q_i and t_i.

The first-order conditions for this constrained maximization are:

$$(3)\quad (1+r)^{-i}\{p_i+t_i(p_2-p_1)-C'(q_i)\} = -\lambda_{i-1}+\lambda_i(1-t_i), \qquad i=2,3\ldots,$$

$$(4)\quad (1+r)^{-i}(p_2-p_1)=-\lambda_i, \quad i=1,2\ldots,$$

where the λ_i are Lagrange multipliers.

Substitute (4) in (3) and simplify to obtain the following necessary condition for profit maximization:

$$(5)\quad C'(q_i)=p_1(1+r)-rp_2, \qquad i=2,3\ldots.$$

Thus, in each period, each firm will choose a value of $q_i=q^*$ which satisfies equation (5). In each period (except, perhaps, for the first whenever $q_1\neq q^*$), t_i will equal M/q^*.

The optimality condition (5) is readily interpretable. To increase by one the number of customers who will visit the shop in each future period, the firm must charge some current customer p_1 rather than p_2, and continue charging him p_1 in all future periods. The firm's loss of revenues is then p_2-p_1, and the discounted value of future profits from this added customer is $[p_1-C'(q)]/r$. Such an addition is worthwhile if and only if $(p_2-p_1)<[p_1-C'(q)]/r$, which condition leads to equation (5).

It may be profitable for a firm to charge all of its customers a price of p_2; this will occur if $p_1(1+r)-rp_2$ is negative. Otherwise, the firm maximizes profits by charging a fraction t of its customers p_2 and charging the others p_1. Observe that a firm's expense in performing a particular repair is irrelevant in determining the price to charge for that repair. That is, there is no reason for the firm to charge a higher price for difficult repairs than for simple ones. The firm may, if it chooses, adopt a rule of thumb according to which it applies the low price only for simple repairs; such a policy may be especially desirable if consumers have some notion of the type of repair required in a particular instance. But the central point remains that in markets with imperfect consumer information, there is no necessary relation between the price of a service and the firm's cost of providing it.

The discount rate relevant to the firm's decision reflects the length of time between successive appearances of repeat customers. The longer the interval between visits of a customer, the longer the firm must wait to reap the benefits of repeat patronage. Other things equal, therefore, a firm will overcharge summer visitors to a town more frequently than it overcharges permanent residents, and overcharge purchasers of (infrequent) roof repairs more frequently than purchasers of plumbing services.

The endogenous variables in the model presented so far are q (the steady-state number of customers at each shop), f (the fraction of all customers charged p_2),[2] and N (the number of shops in the market). These values must represent not only the end result of a firm's decisions, but also be consistent with the conditions for market equilibrium. I assume that each firm maximizes profits and that the number of firms is sufficiently large so that strategic considerations can be ignored; the market equilibrium is a Nash equilibrium.

The equilibrium conditions are that, in each period, each firm has the number of

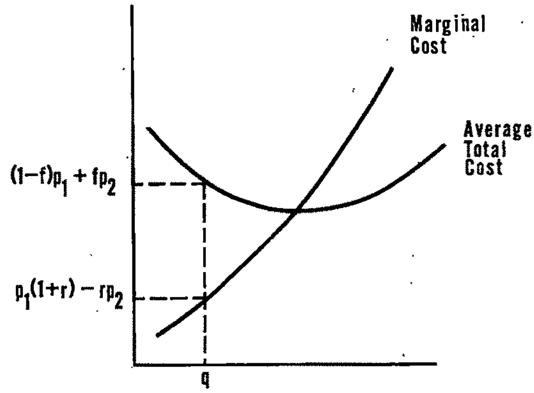


FIGURE 1

customers it desires so that it has no incentive to alter the fraction of customers it charges p_2; each firm earns zero economic profits; and the number of customers in each shop is equal to the total number of customers in the market divided by the number of shops. The equations corresponding to these conditions are for the first, equation (5) above; the average price is equal to average total cost, or,

$$(1-f)p_1+fp_2=C(q)/q \tag{6}$$

and

$$q=Q/N. \tag{7}$$

These equilibrium values are depicted in Figure 1. Although corner solutions are possible, I shall discuss only the nontrivial case in which the equilibrium value of f lies between 0 and 1, so that prices will be dispersed in the market (i.e., both p_1 and p_2 are charged).

I will not give a complete exposition of the adjustment mechanism by which equilibrium is attained, but a sketch of the process can be be helpful. Denote equilibrium values by f^e, q^e, and N^e. Each firm believes that a decrease in its choice of t will increase its future sales, though the number of customers at each shop remains unchanged if *all* shops decrease t to the same extent. Suppose that the market is initially out of equilibrium: $q=q^e$, $N=N^e$, but $f>f^e$. In such a situation, each firm earns positive profits and new

[2]The t denotes the proportion of customers at one firm charged p_2, and f the average value of t across all firms in the market.

firms therefore enter the industry, causing a reduction in the value of q for a given level of f. Each firm will decrease t in order to increase the number of customers who visit its shop. With a lower value of f but more shops in the market, firms will incur losses and some of them will leave the industry. An equilibrium in which $f=f^e$, $q=q^e$, and $N=N^e$ can thus be established.

III. Equilibrium Prices

In the preceding sections I treated the variables p_1, p_2, and r as exogenously fixed. This is a reasonable assumption for the last two, but p_1 can be considered an endogenous variable of the model, as is done in this section.

The average price of a repair depends on two factors: the prices p_1 and p_2 that firms charge, and the frequency with which these prices are charged. Consumers wish to go to those shops that charge, on average, the lowest prices. Unfortunately, consumers do not know the prices a firm has charged other customers. They do, however, understand the incentives firms face, and can form rational expectations of the average cost of a repair at some store.

Let the value of p_1 at virtually all stores be p_1^e, and let $(1-f)$ be the proportion of customers charged this price. The expected price of a repair in the market is then $K^e=(1-f)p_1^e+fp_2$. Consider next a particular shop, A. This firm knows that a customer will return to the shop if it charges a price p_1^e, so that it has no incentive to charge a price lower than that. On the other hand, shop A could increase its profits if, instead of charging a price p_1^e to induce a customer to return, it could charge a price p_1^A greater than p_1^e that had the same effect on his behavior. Does such a price exist? A consumer understands the firm's incentives and by means of equations (5), he can calculate the value of t at shop A, t^A, when $p_1=p_1^A$; he can then calculate the expected cost of a repair at this shop, $K=(1-t^A)p_1^A+t^Ap_2$. If $K\leq K^e$ the consumer will return to the shop even though it charged him a price p_1^A which is greater than p_1^e. Thus, p_1^e is an equilibrium value of p_1 only if $dK/dp_1>0$ when evaluated at p_1^e; if this holds no one firm has an incentive to charge a price of p_1 that differs from p_1^e when at all other firms $p_1=p_1^e$.

From the definition of K it follows that

$$(8)\quad dK/dp_1=(1-t)+(p_2-p_1)(dt/dp_1).$$

But dt/dp_1 equals $(dt/dq)(dq/dp_1)$; from the condition that $t=M/q$ we obtain $dt/dq=-M/q^2$, and from equation (5) we find that $dq/dp_1=(1+r)/C''(q)$, so that

$$(9)\quad dt/dp_1=-M(1+r)/[q^2C''(q)].$$

Substitute equation (9) in equation (8) to obtain

$$(10)\quad \frac{dK}{dp_1}=(1-t)-\frac{(p_2-p_1)M(1+r)}{q^2C''(q)}.$$

The meaning of dK/dp_1 should be clear. Suppose that all firms but one charge prices of p_1 and p_2, and that one firm charges prices $p_1+\varepsilon$ (where ε is a small number) and p_2. The value of dK/dp_1 is the amount by which the average charge at this one shop differs from the average charge at all the other shops. Observe that dK/dp_1, may be negative: that is, an *increase* in the minimum charge p_1 may *decrease* the expected cost of a repair. This will occur, for example, when p_1 is zero. For then each firm maximizes profits by always charging p_2, and an increase in p_1 may induce firms to sometimes charge p_1 rather than p_2, thereby decreasing the average charge.

The conditions for p_1^e to be an equilibrium price are that dK/dp_1 (evaluated at p_1^e) is nonnegative, that $t=f$, and that $M=qf$. Making use of equation (10) this means that in equilibrium

$$(11)\quad p_1\geq p_2-(1-f)qC''(q)/[(1+r)f].$$

It is obvious that there may exist more than one equilibrium value of p_1. The most interesting of these is that at which average price K is at a minimum, or that value of p_1 that satisfies

$$(12)\quad p_1=p_2-(1-f)qC''(q)/[(1+r)f].$$

Not only does this value of p_1 minimize the average charge, but it is also reasonable to suppose that if p_1 is initially less than this value, firms will increase their minimum charges so that, in equilibrium, equation (12) is satisfied. For these reasons, *the* equilibrium value of p_1, p_1^e, is henceforth defined as the one that satisfies equation (12). To recapitulate, if at some firm p_1 was greater than p_1^e, the average charge at that firm would be higher than at all other firms. Consumers, therefore, would not return to a shop whose minimum charge differs from p_1^e, given that all other firms adopt a minimum charge of p_1^e. In view of such consumer behavior, a firm's optimal strategy is to adopt a value of p_1 equal to p_1^e when other firms use that value.

An important feature of this result is that the equilibrium value of p_1 is not identical to the socially optimal value of p_1, the value that would minimize the average price of a repair. The reason for this should be clear. For any given values of p_1 and p_2, a firm's optimal choice of t is a function of the relative frequency f, with which all *other* firms overcharge. The socially optimally value of p_1, however, must consider the effect df/dp_1 rather than dt/dp_1; individual consumers and firms only consider the latter term.

To proceed more analytically, consider df/dp_1, the effect of a small increase in the value of p_1 at *all* firms on the frequency with which firms charge p_2. Take the total derivative of equations (5) and (6), that represent the equilibrium conditions, to obtain

$$(13)\quad \begin{bmatrix} C''(q) & 0 \\ \dfrac{qC'(q)-C(q)}{q^2} & -(p_2-p_1) \end{bmatrix} \times \begin{bmatrix} \dfrac{dq}{dp_1} \\ \dfrac{df}{dp_1} \end{bmatrix} = \begin{bmatrix} 1+r \\ 1-f \end{bmatrix}.$$

The solutions to this system of equations are

$$(14)\quad dq/dp_1 = (1+r)/C''(q);$$

$$(15)\quad df/dp_1 = \frac{-q^2C''(q)(1-f)+(1+r)[qC'(q)-C(q)]}{q^2C''(q)(p_2-p_1)}.$$

The second-order condition for profit maximization requires that marginal cost be increasing, so that the denominator of equation (13) is positive. In equilibrium, average total cost is greater than marginal cost so that the term $(1+r)[qC'(q)-C(q)]$ is negative. It follows that the right-hand side of (15) is negative and an increase in p_1 for all firms will decrease the equilibrium frequency of overcharging.

I demonstrate next that $dt/dp_1 > df/dp_1$, or, making use of equations (9) and (15), that

$$(16)\quad -fq(1+r)/[q^2C''(q)] > \frac{-q^2C''(q)(1-f)+(1+r)[qC'(q)-C(q)]}{q^2C''(q)(p_2-p_1)}.$$

Multiply both sides of (16) by $-(p_2-p_1)/q$, use equations (5) and (6), and simplify to obtain

$$f(1+r)(p_2-p_1)/q^2C''(q) < [qC''(q)(1-f)+(1+r)r(p_2-p_1) + (1+r)f(p_2-p_1)]/q^2C''(q),$$

or

$$(17)\quad \frac{qC''(q)(1-f)+(1+r)r(p_2-p_1)}{q^2C''(q)} > 0.$$

As before, $C''(q)$ must be positive, and by assumption p_2 is greater than p_1. Inequality (17) therefore always holds, which proves that $dt/dp_1 > df/dp_1$. Recall that both dt/dp_1 and df/dp_1 are negative. The result

therefore means that the frequency of over-charging will decrease more if p_1 increases for *all* shops than if it increases for only one shop. It follows immediately that the equilibrium "just" price is smaller than the optimal just price; a policy that increased p_1 above its equilibrium level would lead to a decline in the average price of a repair.

IV. Conclusion

Economists usually view pronouncements concerning the "proper" or "normal" price that should be charged for a service as a mechanism to facilitate price fixing. The practices of local medical societies and of bar associations to set guideline fees have been condemned on these grounds. Yet it is somewhat surprising that the associations have issued such guidelines. The ban on advertising greatly impedes comparison shopping, and the physician or lawyer therefore possesses a large degree of monopoly power over a customer who arrives at his or her office. It appears that these professionals could charge the profit-maximizing monopoly price without regard to the prices charged by other firms, or to the guidelines supposedly set by the price-fixing authority.

In this model, however, such guidelines serve not necessarily as the agreed upon price the firm should charge, but as a pronouncement of the just price. An interesting possibility arises that such pronouncements may serve a socially useful purpose. I argued that the equilibrium just price is not identical to the optimal just price. The pronouncement of a just price by an authoritative body, a price that is adopted by most firms as the level of p_1, can serve to *increase* consumer welfare.

REFERENCES

Darby, Michael K. and Karni, Edi, "Free Competition and the Optimal Amount of Fraud," *Journal of Law and Economics*, April 1973, *16*, 67–88.

Diamond, Peter A., "A Model of Price Adjustment," *Journal of Economic Theory*, June 1971, *3*, 156–68.

Glazer, Amihai and Hassin, Refael, "The Economics of Cheating on the Taxi Market," *Transportation Research*, 1983, *17A*, 25–31.

Phelps, Edmund S. and Winter, Sidney G., "Optimal Price Policy under Atomistic Competition," in Edmund Phelps, ed., *Microeconomic Foundations of Employment and Inflation Theory*, New York: W. W. Norton, 1972.

Rothschild, Michael, "Models of Market Organization with Imperfect Information: A Survey," *Journal of Political Economy*, November/December 1973, *81*, 1283–308.

Smallwood, Dennis E. and Conlisk, John, "Product Quality in Markets Where Consumers are Imperfectly Informed," *Quarterly Journal of Economics*, February 1979, *93*, 1–23.

Cost-Benefit Analysis under Uncertainty: Comment

By Robert Mendelsohn and William J. Strang*

Decision makers performing cost-benefit analysis must often deal with the problem of how to aggregate the benefits across states of nature accruing from an uncertain public investment project.[1] Option price and the expected value of consumer's surplus are two potential measures of these aggregate benefits.[2] The expected value of surplus has been proposed because it is readily measured and because risk pooling (Paul Samuelson, William Vickrey; 1964) and risk spreading (Kenneth Arrow and Robert Lind, 1970) tend to encourage risk neutral behavior. Option price has been favored on the vague notion that people would be willing to pay something extra above expected surplus to preserve the opportunity to purchase a good (Burton Weisbrod, 1964). As Daniel Graham cogently argues in this *Review* (1981), however, option price is but one of an infinite number of contingent payment schemes. The literature has provided no justification for focusing upon it as an ideal measure of benefits under individual risk.

Graham further argues that policymakers ought to adopt the compensating contingent payment plan which maximizes expected revenue. This maximum payment plan, by definition, is never less and will often exceed any other contingent payment scheme. Consequently, Graham argues that both the expected value of surplus and option price are underestimates of the true value of project benefits.

Our purpose in this comment is twofold: first, we show the role of project and nonproject insurance in a model of individual risk; and, second, we argue that option price, not the maximum payment plan, is the optimal rule when no fair insurance is available. In Section II, we show that if fair insurance is available against all risks, all contingent payment plans yield identical revenue. A similar result holds if insurance is available for nonproject risks and the effect of the project to an individual is small (the Arrow-Lind model). In Section III, we explore the case where either the project has a large uninsurable effect on the individual, or there is no insurance against even nonproject risks. We argue that the very phenomena (moral hazard, adverse selection, and complexity) that eliminate the market for private insurance also prevent the government from making otherwise desirable contingent payments. If contingent payments are too costly, the government's only remaining choice is to collect payments that are constant across states, which makes option price the relevant measure of benefits.

I. Insurance and the Evaluation of Benefits

Let us begin with a model of a representative consumer. We assume the individual's expected utility function without the project can be written

$$\bar{V} = \sum_i \Pi_i V_i(M_i, 0), \tag{1}$$

*Departments of Economics, University of Washington, Seattle, WA 98195, and Washington University, St. Louis, MO 63130, respectively. We thank Myrick Freeman, Dick Hartman, Richard Schmalensee, and Daniel Graham for helpful comments. This research was partially supported by a grant for natural resource regulation from the Sloan Foundation.

[1] In this comment, we confine our attention to problems of individual risk and certain costs. The appropriate measure of benefits with collective risk, where everyone shares each state of nature together, is defined by a simple budget constraint. Benefits should be measured as costs are distributed. As Graham (1981) notes, when costs are certain, collective risk benefits should be measured by option price.

[2] Option price is the payment individuals would make for the project independent of the state of nature. The consumer's surplus payment is what the individual is willing to pay for the project in each state of nature. Note that the expected value of surplus is a weighted average of what people would pay across all the possible states of nature. In contrast, demand studies, although unbiased, frequently measure the value of surplus at one moment in time (one state of nature).

where Π_i is the probability and M_i is the endowment in state of nature i and 0 denotes no project. We define a contingent payment plan as the set of payments γ_i in each state of nature an individual will make with the project that leaves the individual at his initial (no project) level of utility. When insurance is not available:

$$\sum_i \Pi_i V_i(M_i - \gamma_i, 1) = \overline{V}, \tag{2}$$

where 1 denotes the project is in place. Note that this use of compensating payment implies that only users will actually pay for a project. In general, nonusers pay for public projects and this introduces the traditional problems of comparing utilities across individuals. Although a potential compensation criterion may, in fact, solve this problem, this issue is beyond the scope of this note.

As Graham notes, there is an entire locus of these payment plans which he labels the willingness-to-pay (*WTP*) frontier. Taking the total differential of (2) and rearranging, we obtain the slope of the *WTP* locus:

$$\frac{d\gamma_j}{d\gamma_i} = -\frac{\Pi_i V_i'(M_i - \gamma_i, 1)}{\Pi_j V_j'(M_j - \gamma_j, 1)} \qquad \forall_{i,j}, \tag{3}$$

where V_K' is the marginal utility of income. Because of diminishing marginal utility of income, the *WTP* locus is convex to the origin (see Graham's Figure 1, p. 717).

The maximum payment plan is the contingent payment scheme which maximizes expected revenue ($\Sigma \Pi_i \gamma_i$). Assuming the necessary second-order conditions and maximizing expected revenue subject to (2) yields the following first-order conditions:

$$V_j'(M_i - \gamma_j, 1) = V_i'(M_i - \gamma_i, 1) \qquad \forall_{i,j}. \tag{4}$$

The maximum payment scheme equates the marginal utility of income across all states of nature. All other schemes yield less expected revenue because of the concavity of the *WTP* locus.

Let us now examine the effect of introducing fair insurance in this model. Fair insurance allows the individual to reallocate expected dollars across states of nature. The individual gains some additional flexibility with no change in expected income:

$$\sum \Pi_i I_i = 0, \tag{5}$$

where insurance payments I_i can be positive or negative. Given each contingent payment scheme, the individual allocates insurance payments to maximize his expected utility subject to (5) where expected utility is now:

$$\sum \Pi_i V_i(M_i - \gamma_i - I_i, 1). \tag{6}$$

Differentiating (6) with respect to I_i subject to (5) yields the following first-order conditions:

$$V_i'(M_i - \gamma_i - I_i, 1) = V_j'(M_j - \gamma_j - I_j, 1) \qquad \forall_{i,j}. \tag{7}$$

Fair insurance allows individuals to equate their marginal utilities of income across states regardless of the initial distribution of their incomes. Because the marginal utilities of income are equated for all the contingent payment plans, the slope of the *WTP* locus (3) reduces to

$$d\gamma_j / d\gamma_i = -\Pi_i / \Pi_j \qquad \forall_{i,j}, \tag{8}$$

which is just a straight line. With insurance, all the contingent payment schemes yield the same expected payments.[3] Expected payments between one contingent payment scheme and another differ only when they provide different amounts of insurance.

Both project and nonproject risks are relevant in this analysis. Suppose insurance is available for nonproject risks. Without the project, people equate their marginal utilities of income across states. Now suppose be-

[3] For example, option price and expected surplus yield the same expected payment when fair insurance is available, so option value is zero. This specific result was first proven by Richard Schmalensee (1972).

cause of risk spreading, the individual risk of the project is small. That is, suppose the project barely alters the marginal utility of income across states. If the marginal utility of income is equal across states, aggregate *WTP* locus (3) remains a straight line. All contingent payment plans yield the same expected payment equal to the expected value of consumer's surplus. This is essentially the Arrow-Lind argument for using the expected value of project benefits as the measure of true benefits.

The model first explored by Schmalensee and expanded by Graham includes the case where nonproject risks cannot be insured against. In this case, even if projects are small (because of risk spreading), the marginal utilities of income across states are not equated. Different contingent payment schemes yield different revenues even when the individual benefits of a project are too small to alter the marginal utility of income in any state. Contingent payments can yield different revenue not only because of their ability to insure against project risks, but also because of their ability to insure against nonproject risks.

II. Optimal Contingent Payment Plans

If contingent claims markets exist, the expected value of consumer's surplus is the appropriate measure of benefits. As Graham argues, these claims markets (private insurance) may not exist. This may be due to adverse selection (asymmetric information about the probability of occurrence of states of nature), complexity (too many states of nature to inexpensively develop a payment scheme), or moral hazard (the insurance itself alters the distribution of states of nature). In the absence of contingent claims markets, Graham argues for the use of the maximum payment plan as a measure of benefits. In this section, we discuss our reservations about this recommendation.

In order for the government to collect more revenue than the sum generated by the expected value of surplus, the government must provide an additional service. By collecting more funds in states of nature where the marginal utility of income is low and less funds when the marginal utility of income is high, contingent payment schemes could help individuals equate their marginal utility of income across states. The way the payments are collected could actually provide insurance to taxpayers. For this extra service, the taxpayer would be willing to pay more expected revenue. Although Graham is correct that one would prefer contingent payment schemes that provide more revenue, it does not follow that the correct measure of benefits is the maximum revenue contingent payment scheme and the other measures are underestimates. First, if the government is able to collect contingent payments which provide insurance, why doesn't it engage in this activity without public projects? The provision of fair insurance against individual risk would clearly make all citizens better off even if no projects are undertaken. With the government independently providing fair insurance, all contingent payment schemes for projects would yield identical revenues.

Second, in the absence of contingent claims markets, the government's ability to collect revenue contingent on the state of nature will be severely restricted. Contingent claims markets are assumed not to exist because private insurers are unable, at reasonable cost, to devise an instrument that will make or collect payments contingent on the true state of nature. We question whether governments are better suited to construct such instruments than private firms specializing in this business. If moral hazard, adverse selection, and complexity prevent private firms from making or collecting contingent payments, it is also unlikely that the government could make or collect such payments. The very assumptions that eliminate the market for insurance are likely to prevent the government from providing insurance as well. If it is too costly to collect revenue contingent on the state of nature, the government will have to choose revenue schemes which are independent of the state of nature. As Graham argues, the benefits should be measured using the contingent payment scheme actually used. In this case, because the government actually requires payments independent of the state of nature, option price is the relevant measure of benefits.

III. Conclusion

When fair insurance is available against all risk, or when nonproject risks are insurable and the effect of projects on individuals are small, all contingent payment schemes yield identical revenue. The expected value of surplus is an appropriate measure of value in this case.

When no insurance is possible, or when project insurance is not possible and projects have large impacts on individuals, each contingent payment scheme will potentially yield different revenues. The very phenomena that eliminate insurance, however, will also prevent the government from collecting contingent payments. Collections independent of the state of nature may be the government's only feasible alternative, and, consequently, option price would be the appropriate measure of benefits.

Without adequate insurance, the government may be in the awkward position of observing the expected value of surplus measure of benefits, but desiring the option price measure of benefits. When the sign of option value, the difference in revenue between option price and the expected value of surplus, is known, it is clear that the observable expected value of surplus measure should be adjusted. For example, if the uncertainty is solely due to variations in income across states of nature and the publicly supplied good is normal, option value is negative.[4] Further, the magnitude of this effect may be large (see A. Myrick Freeman, 1984).

When uncertainty is related to price variations, however, the sign of option value is ambiguous without further assumptions.[5] As Schmalensee argues, if the sign of option value is ambiguous, no bias is introduced by using the expected value of surplus to evaluate the benefits of public projects. Thus, we find in most practical circumstances that the appropriate measure of the benefits of uncertain projects is the expected value of consumer's surplus.

[4]As income increases, the marginal utility of income falls and the benefits increase (when the good is normal). The expected value of surplus, that collects payments depending on benefits, will collect more benefits in states of nature with lower marginal utility of income than will option price. Option value is negative in this case.

[5]Many of the papers on option value have incorrectly argued that it is positive. The more formal models, Schmalensee, Graham, and Richard Hartman and Mark Plummer (1982), recognize that option value is ambiguous in general. See V. Kerry Smith (1983) for an able review of the literature.

REFERENCES

Arrow, Kenneth J. and Lind, Robert C., "Uncertainty and the Evaluation of Public Investment Decisions," *American Economic Review*, June 1970, *60*, 364–78.

Cichetti, Charles and Freeman III, A. Myrick, "Option Demand and Consumer Surplus: Further Comment," *Quarterly Journal of Economics*, August 1971, *85*, 528–39.

Freeman, A. Myrick III, "The Sign and Size of Option Value," *Land Economics*, February 1984, *60*, 1–13.

Graham, Daniel A., "Cost-Benefit Analysis under Uncertainty," *American Economic Review*, September 1981, *71*, 715–25.

Hartman, Richard and Plummer, Mark, "Option Value Under Income and Price Uncertainty," unpublished, 1982.

Krutilla et al., John, "Observations on the Economics of Irreplaceable Assets," in Allen V. Kneese and Blair T. Bower, eds., *Environmental Quality Analysis: Theory and Method in the Social Sciences*, Baltimore: Johns Hopkins University Press, 1972.

Samuelson, Paul A., "Principles of Efficiency: Discussion," *American Economic Review Proceedings*, May 1964, *54*, 93–96.

Schmalensee, Richard, "Option Demand and Consumer's Surplus: Valuing Price Changes Under Uncertainty," *American Economic Review*, December 1972, *62*, 813–24.

Smith, V. Kerry, "Option Value: A Conceptual Overview," *Southern Economic Journal*, January 1983, *49*, 659–68.

Vickrey, William, "Principles of Efficiency: Discussion," *American Economic Review Proceedings*, May 1964, *54*, 88–92.

Weisbrod, Burton, "Collective Consumption Services of Individual Consumption Goods," *Quarterly Journal of Economics*, August 1964, *77*, 71–77.

Cost-Benefit Analysis under Uncertainty: Reply

By DANIEL A. GRAHAM*

Robert Mendelsohn and William Strang raise a number of issues which require clarification. Their misrepresentations of my results in several cases are easily corrected. The analytical flaws in their own arguments are more difficult.

Turning first to the issue of misrepresentations, they incorrectly assert that I recommended the use of the maximum expected value of alternative contingent payment schemes as the measure of project benefits. Even a cursory reading of my introduction reveals that option price, which does not maximize the expected value of revenue, is the appropriate measure of benefit in situations involving similar individuals, collective risk and certain costs. Despite their claims to the contrary, I have not suggested that option price is an "underestimate" of the true value of project benefits—option price is precisely the true value of project benefits *in this case*. Contradicting themselves in their first footnote, they assert that I recommend the use of option price for cases involving certain costs and collective risks, thus ignoring the crucial role of similar preferences. Other examples could be mentioned, but it is perhaps sufficient to suggest that any reader wishing to know what I actually said would do well to review my paper.

The appropriate measure of benefits is the subject of the third section of my paper. No mention is made in this general argument of expected value calculations, similar preferences, certain costs, or any of the other special cases that are considered in the fourth section of my paper. These special cases are exactly that—situations in which the general answer, because of the special circumstances, takes a relatively simple form. It is, of course, crucial not to forget the special circumstances when considering these results.

*Professor of Economics, Duke University, Durham, NC 27706.

Although Mendelsohn and Strang do not mention the distinction between collective and individual risks save for a passing reference in their first footnote, this issue is central to the consideration of the appropriate measure of project benefits. While the general procedure developed in my third section is compatible with *any* joint distribution of individual states, option price measures the benefit to similar individuals of a project involving a collective risk when cost is certain. Expected value calculations, on the other hand, are only appropriate in situations involving individual risks. These issues were examined in detail in my fourth section. In dealing with the special case involving individual risks, I argued that the magnitude of aggregate payments would be equal to the sum of the expected values of the individual payments with virtual certainty. This was hardly a profound observation, but it led to the result that if contingent payments are to be collected in such circumstances, one should consider those payments that have the largest expected value since all contingent payment schemes, including that associated with "expected surplus," involve the same difficulties of adverse selection, moral hazard, and so forth. Mendelsohn and Strang, in expressing the view that "...for the government to collect more revenue than the sum generated by the expected value of surplus, the government must provide an additional service" (p. 1098), fail to recognize that collecting the wrong contingent payments—those associated with the surplus point—involve the same difficulties as collecting the right contingent payments—those associated with the fair bet point.

This mistake is not eliminated by their argument that when fair insurance is available, the individual's willingness to pay locus becomes linear. The presence of such a market makes it possible for an individual to "trade" a given payment "liability" in the market for contingent claims for another,

less onerous liability with the same expected (market) value. Such a market makes it possible for the individual to "consume" contingent claims corresponding to the fair bet point while making *any* contingent payments with the same expected value as the fair bet point. This would allow the individual, for example, to make sure payments while consuming contingent claims corresponding to the fair bet point, thereby eliminating the necessity for the government to collect contingent payments. This indifference to payments with the same expected value does not, however, alter the fact that the expected value of these payments is equal to the expected value of the fair bet point and *not* the expected value of surplus. Recall, in this regard, that "surplus" in a particular state is defined as the largest payment the individual would be willing to make for the project *if that state were certain*. The presence or absence of contingent claim markets is irrelevant to this definition and thus to the calculation of expected surplus. The conclusion of their Section I should, therefore, be corrected to read: a project that involves independent risks for which actuarially fair insurance markets exist, has benefits equal to the sum of the expected values of the *fair bet points* of the affected individuals and these expected values can be collected as sure payments.

Even the correct version of their result is of dubious interest since it ignores the question of whether existing markets can be expected to provide opportunities for insuring risks which do not currently exist and will not exist in the future unless the project is undertaken. The standard assumption is that projects involve risks that are independent of all existing risks—otherwise the measurement of project benefits would involve portfolio-type considerations regarding the contribution of the project to the joint distribution of all risks. Can existing markets provide actuarially fair opportunities for insuring risks that are independent of the risks for which these markets were created? Mendelsohn and Strang do not examine this issue.

Their proposition that projects which entail individual risks and "small" changes in the marginal utility of income across states give rise to linear *WTP* loci is also incorrect. Suppose, for example, that

$$V_i(\ ) = \log_e(M - y_i); \qquad i = 1,2,$$

where there is no risk other than the project and the marginal utility of income across states is not affected *at all* by the project. Supposing that states 1 and 2 are equally likely one obtains the *WTP* locus

$$(M - y_1)(M - y_2) = \text{constant},$$

which is certainly not linear. Considering first the case of individual risks confronted by 2 identical individuals, we have 3 "social states": both experience state 1; one person experiences state 1 and the other person experiences state 2; and both experience state 2. The aggregate *WTP* locus would then consist of triples of the form $(2y_1, y_1 + y_2, 2y_2)$ where y_1 and y_2 belong to the locus identified above—again not linear. Since Mendelsohn and Strang are discussing individual risks, their appeal to the Arrow-Lind result, which involves a collective risk, is incorrect. On the other hand, with N identical individuals facing a *collective risk*, the aggregate *WTP* locus would be

$$(M - y_1/N)(M - y_2/N) = \text{constant}.$$

Although not linear, this curve becomes linear in the limit as N approaches infinity—the Arrow-Lind result. The correct conclusion in this case would read: a project that entails a collective risk for an arbitrarily large number of similar individuals gives rise to a *WTP* locus that is approximately linear. Whether or not the slope of this *WTP* locus is equal to the "fair odds" depends upon the effect of the project upon relative marginal utilities of income across states. Again, even the correct version is of limited interest given the simpler proposition for the same case presented in my earlier paper: the value of the project is N times the option price of the representative individual.

Mendelsohn and Strang correctly point out that the presence or absence of contingent claims markets has important implications for benefit measurement, but express reserva-

tions with regard to the government use of contingent payment schemes. Why, they ask, if the government is able to collect contingent payments, doesn't it engage in this activity to provide insurance in the absence of the project? Insurance against what? The idea of a project with uncertain benefits, presumably, is that the provision of such a project with option price or "sure" financing would expose individuals to a new uncertainty or risk—the project will be worth more, net of this payment, in some states than others. Remove the project and you remove the purpose of the insurance.

The example I gave of a contingent payment mechanism for Yellowstone Park that involves a combination of sure taxes and visitor charges is an example of such insurance—the individual pays more in states in which the park is worth more, that is, when he visits. Even if visitor charges for a nonexistent park were conceivable, what purpose would they serve? The use of contingent payment mechanisms is not limited, moreover, to insurance for independent risks. My example of the financing of the dam, whose benefits depend upon the collective risk of whether the year is wet or dry, with a combination of sure taxes and water charges is an example of a contingent payment mechanism for a collective risk—farmers pay more in dry years when the dam is more valuable, that is, when they use more water.

A second objection to the use of contingent payment mechanisms raised by Mendelsohn and Strang concerns the difficulties associated with collecting payments contingent upon the "true state of nature." They do not take into consideration that *all* payment schemes involve a specification of payments to be collected in each true state of nature. The only difference between a contingent payment mechanism and a "certain" payment mechanism is that the former involves possibly different payments in different states while the latter requires the same payment in every state. The examples of contingent payments for Yellowstone Park and the dam are easier to imagine than a tax which collects the same payment from an individual no matter which state occurs. Taxes that do not depend upon the individual's income, the market value of his property, the amount of gasoline he consumes, or any other variable that is state dependent are as difficult to imagine as the proverbial "lump sum" tax. The trick, of course, is to collect the right payment in every state, but this may well be no more difficult than collecting the same payment in every state.

REFERENCES

Arrow, Kenneth J. and Lind, Robert C., "Uncertainty and the Evaluation of Public Investment Decisions," *American Economic Review*, June 1970, *60*, 364–78.

Graham, Daniel A., "Cost-Benefit Analysis under Uncertainty," *American Economic Review*, September 1981, *71*, 715–25.

Mendelsohn, Robert and Strang, William, "Cost-Benefit Analysis under Uncertainty: Comment," *American Economic Review*, December 1984, *74*, 1096–99.

Labor Supply and Tax Rates: Comment

By JAMES A. WILDE*

In their recent contribution to this *Review* (1983), James Gwartney and Richard Stroup remind us of both the care that must be taken in moving from individual to aggregate impacts and the importance of public good provision in labor supply decisions. Unfortunately their formulation of the interrelationship between tax rates and labor supply leads to incorrect conclusions. In the process of outlining an appropriate illustration, I shall also demonstrate that the so-called Laffer curve may take on a variety of forms.

I. Tax Rates and Optimizing Behavior

As the present supply-side tax cuts shift our attention from equity to efficiency concerns, we clearly must realize that such cuts may produce different results for individuals than for the economy as a whole. No longer do we think of shifting tax burdens from one citizen or group to another, with total tax revenues and public good spending unchanged. Instead, a tax cut for one individual will be accompanied by a similar change for all others. Thus public good provision cannot be assumed constant.

Gwartney and Stroup see this difference as allowing them to assume that the "alleged income effect of the tax cut simply disappears" (p. 448) because the technical production possibilities and the aggregate real income cannot be increased. Since the focus of their model is on the desired labor supply and how it may be affected by tax rates, it seems inappropriate to base this model on a premise of constant aggregate production and income. Consideration should instead be given to the possibility that the full employment level of national income and its associated utility level will be a function of the tax rates under consideration.

*University of North Carolina, Chapel Hill, NC 27514.

Let us follow Gwartney-Stroup in assuming a mixed economy of homogeneous individuals, such homogeneity presumably pertaining to both the abilities and the tastes of those citizens. I center my analysis around a representative individual whose welfare is derived from the consumption of three "goods": leisure (L), a private good (X), and a public good (Z). Merely for illustrative purposes, I assume the following utility function:

$$(1)\quad U = 200X - X^2 + 216Z - Z^2 + 416L - L^2.$$

This specific formulation follows the normal characteristics usually assumed for such goods.

Optimality for this individual can be ascertained by a maximization of his or her utility function, subject to the appropriate budget constraint, which includes the exogenous wage rate (w), price of the private good (p_x), and price of the public good (p_z). In a Robinson Crusoe-type world, the individual will bear the full price of the public good. If we specify $w=4$, $p_x=2$ and $p_z=5$, the weekly budget constraint

$$(2)\qquad (168-L)\cdot w = p_x\cdot X + p_z\cdot Z$$

yields the optimal levels: $L=128$, $X=60$, and $Z=8$, thus making public good spending comprise 25 percent of national income.

Now permit $(n-1)$ other identical individuals to enter this society. These new residents are rivals in good X and nonrivals in good Z. How do these identical individuals select their optimal levels of L, X, and Z? Because of the homogeneity assumption, each must expect others to behave as himself or herself, or be seen as subject to some sort of fiscal illusion. In the absence of such illusion, the individual citizen would ask the government to provide an amount of public goods indicated by the above maximizing process.

TABLE 1—OPTIMAL RESPONSES TO TAX RATES[a]

t	Z	L	X	Income	Utility
0	0	136.0	64.0	128.0	46,784
.1	2.8	133.0	62.9	140.0	46,860
.2	6.1	129.7	61.3	153.2	46,916
.24	7.6	128.4	60.2	158.4	46,927.6
.25	8.0	128.0	60.0	160.0	46,928
.26	8.4	127.7	59.6	161.2	46,927.6
.3	10.0	126.2	58.5	167.2	46,911
.4	14.4	123.0	54.0	180.0	46,826
.5	19.1	120.2	47.8	191.2	46,591
.6	23.6	118.9	39.3	196.4	46,181
.7	27.0	119.8	28.9	192.8	45,533
.8	28.2	124.0	17.6	176.0	44,714
.9	26.7	130.9	7.4	148.4	43,799
1.0	22.6	139.7	0	113.2	42,970

[a]Based on $U = 200X - X^2 + 216Z - Z^2 + 416L - L^2$; $W = 4$, $P_x = 2$, $P_z = 5$, and Z as the social good.

There would be unanimity about the optimal level of public goods, the homogeneity assumption having permitted us to avoid possibly thorny public choice questions. Without implying anything about the effect of population size on the price of public goods, and merely for arithmetic convenience, let us assume that the true overall price of Z is $n \cdot p_z$. Each identical individual will thus face an effective public good price of p_z. With the illustrative prices given above, there will be unanimous consent to have 25 percent of national income devoted to the public sector purchase of Z. This solution satisfies the Samuelsonian conditions for public goods and ought to be what Gwartney and Stroup meant by "the initial level of government expenditure [which] is socially efficient" (p. 448).

Taxation may be introduced into this framework by adding the constraint

$$(3) \qquad t(168 - L)w = p_z \cdot Z,$$

where t represents the marginal and average income tax rate. This balanced governmental budget requirement is compatible with the Gwartney-Stroup assumptions of a constant money supply and an unchanged fiscal deficit. By allowing t to vary from zero to 100 percent, we can observe the responses of our representative individual. Note that if $t = 0$, there are no allowed purchases of public good Z. The individual will have an incentive to work only to the extent that the desirability of private good X is sufficient to warrant sacrificing leisure, surely a well-established notion. But also note that if $t = 1$, there are no allowed purchases of private good X. An individual will have an incentive to work to the extent that the desirability of public good Z is sufficient to warrant sacrificing leisure, a result that seems to have been somewhat overlooked.

Table 1 presents the optimal levels of L, X, and Z under several possible values of t for the model outlined in equations (1), (2), and (3) and with the assumed prices. Observe that the individual will do some work at *all* income tax rates between zero and 100 percent. It can also be seen that a tax rate of 25 percent yields the maximum utility level, a result fully in line with the optimal Z level of 8 described above.

However, the patterns of behavior as the tax rate varies on either side of the 25 percent optimal level is entirely different from what was predicted by Gwartney and Stroup. They tell us that:

> Independent of any additional output emanating from the substitution effect, the reduction in tax rates will clearly lead to a decline in tax revenues. For our general equilibrium model, any expansion in expenditures on

> private goods stemming from the alleged income effect of the tax cut will be exactly offset by a decline in expenditures on public good (because of the decline in tax revenues that finance them). When the initial level of government expenditure is socially efficient, then at the margin, the additional utility derived from an expansion in consumption of private goods is exactly offset by the decline in utility associated with the reduction in consumption of public goods. The alleged income effect of the tax cut simply disappears. Only the substitution effect remains and the tax cut will unambiguously increase the quantity supplied of labor. For a tax cut of discrete value, an individual's valuation of the public goods foregone as a result of the income effect of the tax cut will exceed his or her valuation of the additional private goods now available. Real income will fall because government spending is now below the optimal level. Under these circumstances, the negative income effect will reinforce the substitution effect. Unambiguously, the quantity supplied of labor will increase. [p. 448]

A quick glance at Table 1 reveals that a cut in tax rates from optimality at 25 to 20 percent results not in an increase but in a reduction in labor supply.

Consideration of my illustrative case reveals why the unambiguity asserted by Gwartney-Stroup is destroyed. We shall see that this may be due to confusing interpretations of the income and substitution effects. If t is reduced from .25 to .20 without any change in labor supply, there will be a forced substitution of private goods for public goods (with no change in labor income). The appropriateness of the term "income effect" to describe this situation is dubious, though the utility decline resulting from this discrete change is evident. But what is the expected response to this nonoptimal consumption mix? With the ratio of private-to-public-good consumption being determined by the tax rate and thus outside the control of the individual, only variations in labor supply can be used to improve welfare. This citizen must weigh the value of leisure against the value of this nonoptimal mix of public and private goods. In this case utility maximization results in a "substitution effect" of reduced labor supply and thus lower income.

Table 1 also reveals the ambiguity of the labor supply consequences of a tax cut, even when the optimal spending mixes and levels characterize the starting point. Suppose that X had been the public good and Z the private good. Allocative efficiency would have occurred at $t = 75$ percent. A reduction in this tax rate to, say, 70 percent would cause an increase in labor supply from 40 to 41.8 hours per week. Thus we can conclude that a reduction in the tax rate from its optimal public-good level may result in either an increase or a decrease in labor supply, depending on the marginal utility profiles of the goods in question.

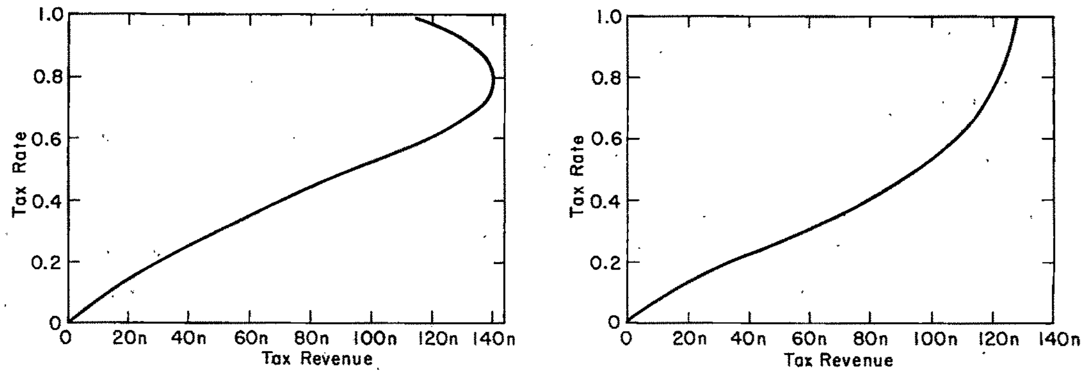
II. Maximum Work, Production, and Tax Revenue

Gwartney and Stroup go on to imply that the maximum labor supply will occur where $t = 0$:

> If we relax our assumption that the initial output of public goods was optimal, the income effect associated with a discrete tax cut might be either positive or negative within our general equilibrium model. We have already pointed out that the income effect will be negative if individuals value the public goods that must be foregone as a result of the tax cut more than the expansion in private goods stemming from their increase in disposable income. In this instance, the income and substitution effects reinforce each other. [p. 449]

Returning to the case of Z as the public good, this would be illustrated by a further tax cut from 20 to 10 percent. Rather than the predicted labor supply increase, we see a reduction all the way to a zero tax rate. Once again we find that the traditional micro concepts of the income and substitution effects cannot be relied on when tax rates and public goods vary simultaneously.

The so-called "Laffer curve" literature has also addressed the association between tax



A. Z AS PUBLIC GOOD B. X AS PUBLIC GOOD[a]

FIGURE 1. LAFFER CURVE

[a]Based on $U = 200X - X^2 + 216Z - Z^2 + 416L - L^2$, $w = 4$, $p_x = 2$, $p_z = 5$, and population n.

rates and labor supply. Jude Wanniski asserts that:

> If the tax rate is zero, people can keep 100 percent of what they produce in the money economy. There is no governmental "wedge" between earnings and after-tax income, and thus no governmental barrier to production. Production is therefore maximized, and the output of the monetary economy is limited only by the desire of workers for leisure. [1979, p. 7]

While clearly there is more to production than merely labor inputs, in a *ceteris paribus* world maximum production will occur where labor supply (totally utilized under a full employment assumption) is maximized. Table 1 shows that this need not be achieved when tax rates are zero.

Wanniski further clouds this issue by maintaining that both maximum production and maximum tax revenue are achieved at "the point at which the electorate desires to be taxed" (p. 8). If this refers to what I have previously called the optimal level of public goods, Table 1 shows that such simultaneous maximization need not take place. Instead we find maximum utility at $t = .25$, maximum production at $t = .60$ and maximum tax revenue at $t = .80$. This illustration thus demonstrates the likely inability of a supply-side tax policy to attain multiple goals (such as production, utility or tax revenues).

In fact, the same utility function (equation (1)) with the same exogenous wage rate and prices yields rather different Laffer curves depending on which of the two commodities is identified as the public good. Figures 1A and 1B show that not only do 100 percent tax rates provide positive tax revenue in these cases, but also that it is possible to generate a monotonic Laffer curve with maximum revenues at $t = 100$ percent.

III. Summary

I have specified a model for understanding the interrelationship between public goods, tax rates, and labor supply. In a supply-side model in which tax rates are reduced for all citizens simultaneously, the amount of public goods must be adjusted and determined if a balanced budget is to be maintained. Thus the model must specify the public choice technique for selecting the level of public goods. If a homogeneous population is assumed, a simple extrapolation of individual choice can be used, with labor supply and the public budget being jointly determined. Such a method was utilized here. The alternative would entail some sort of fiscal illusion, with the individual treating public goods as exogenous. The method by which each labor supplier would estimate this public good supply would have to be specified. This latter requirement was not met in the Gwartney and Stroup article.

I have used a simple quadratic utility function with illustrative prices to yield an optimal public good supply roughly in line with present national income ratios. The following conclusions were reached.

1) That it is not necessarily true that labor supply will increase when tax rates are reduced.

2) That the "income-plus-substitution effect" method of analyzing tax changes breaks down when public goods are allowed or required to vary within a three-good general equilibrium model.

3) That some nontraditional Laffer curves can be derived from a plausible model.

4) That tax rates of 100 percent may yield not only positive tax revenue but also maximum tax revenue.

5) That maximum production is likely to occur not when tax rates are zero, nor where tax revenues or utility are maximized.

6) That tax revenues need not be maximized where citizens have attained maximum utility.

The model used here makes no attempt to analyze the macro effects of tax cuts or rates. Nor does it attempt to explore incentives for tax evasion through the barter economy or for tax avoidance through home production. These latter types of activity may be major aspects of actual labor supply decisions with considerable impact on the tax revenue situation. I shall leave for future work the integration of all three aspects of tax response.

REFERENCES

Gwartney, James and Stroup, Richard, "Labor Supply and Tax Rates: A Correction of the Record," *American Economic Review*, June 1983, *73*, 446–51.

Wanniski, Jude, "Taxes, Revenues, and the 'Laffer Curve'," in Arthur B. Laffer and Jan P. Seymour, eds., *The Economics of the Tax Revolt*, New York: Harcourt Brace Jovanovich, 1979.

Labor Supply and Tax Rates: Reply

By JAMES GWARTNEY AND RICHARD STROUP*

In our earlier paper we focused on the internal inconsistency involved in generalizing the impact of a tax change on the individual's work-leisure choice to the aggregate economy. The work-leisure indifference curve analysis for the individual indicates that a tax reduction (increase) increases (reduces) the real income of the individual *by the amount of the tax cut* (increase). However, in aggregate, this will not be true if the government uses marginal tax revenues for useful purposes. If tax revenues are simply rebated in the form of income transfers, a tax reduction (increase) leads to an equal reduction (increase) in income transfers. In aggregate, the change in income transfers will offset the change in disposable income, causing the income effect implied by the individual work-leisure analysis to dissipate. Similarly, if tax revenues are used to supply public sector goods equal in value to the private goods foregone by taxpayers, a change in tax revenues will lead to an offsetting change in the supply of public sector goods. Once again, the income effect of the traditional analysis dissipates. Thus, in a general equilibrium setting, moving from the individual work-leisure analysis to the aggregate involves an internal inconsistency. This was the message of our earlier paper.

We appreciate the interest of James Wilde in this topic. While his approach differs from the traditional work-leisure analysis for the individual, we believe that it also suffers from an internal inconsistency. The central problem with Wilde's approach is his failure to integrate the theory of the free rider into his analysis.[1] In fact, Wilde assumes away the free-rider problem. Therefore, his analysis is inapplicable to large number cases because, under such circumstances, individual behavior will differ from the postulates of his model.

As formulated, Wilde's model is not really an n-person model. The supply of the public good is not determined by the labor supply of n individuals. In fact, there is no reference to the number of individuals of the economy in any of his equations. All individuals are assumed to be identical, so that each can increase the "public good" in proportion to his or her "tax" contribution. Wilde's formulation is for a representative individual who maximizes his or her utility subject to (a) a fixed wage rate, (b) fixed prices for two private goods, and (c) a so-called "tax rate" which simply alters the price ratio (the individual's ability to transform work into each of the goods) between the two goods. What Wilde calls a tax rate is simply an added constraint: a forced division between the two goods purchased by the individual. When that tax rate is low, a high percentage of the wages derived from work are forced into the purchase of x (and a low percentage to z, Wilde's public good). In contrast when the tax rate is high, a high percentage of the wages from additional work effort go to the purchase of z (and little to x). At a tax rate of zero, leisure can be transformed into only good x. At a tax rate of 100 percent, leisure can be transformed only into good z.

The data of Wilde's Table 1 permits the work-leisure tradeoff *of the individual* to determine the total quantity of good z available to him, and thus presumably the quantity available to the entire economy. Clearly, this would not be the case for a multiperson economy. Given the specification of the author, the model simply collapses into a work-leisure, good x/good y problem *for an individual.*

In Wilde's model, individuals "face an effective public good price of p_z" even though "the true overall price of Z is $n \cdot p_z$" (p. 1104). In a multiperson world, this assump-

*Professors of Economics, Florida State University, Tallahassee, FL 32306, and Montana State University, Bozeman, MT 59715, respectively.

[1] See Edgar Browning and Jacquelene Browning (1983) for an excellent discussion of the free-rider problem.

tion implies that when one individual works to supply more of the public good relative to the private good, *all others follow suit*. As the result of this assumption, each *individual* can transform work into additional units of the public good at a price of p_z, rather than its actual price, np_z. In the large number case, this assumption is clearly false. Individuals will recognize that the supply of the public good is largely invariant to their individual work effort. The private-good/public-good price ratio confronted by each individual will be px/np_z, not px/p_z as Wilde assumes. Thus, the incentive of each individual to work to supply the public good will differ substantially from that indicated by Wilde.

One might perceive that Wilde's model is rescued by the assumption of homogeneous individuals. This is not the case. Even with the optimal tax rate, each individual would still face the full price (np_z) of personal additions to the public good. In the large number case, as the free-rider analysis explains, individuals would have little incentive to work more so they could consume more of the public good. For an economy of 200 million, as in the case of the United States, it makes no sense for a taxpayer to work harder in order to generate tax revenues necessary to finance a larger supply of a public good; national defense, for example. Neither does it make any sense to assume that an individual can transform work into the public good at a price p_z because everybody else will supply more public goods (or tax revenues) if the one individual does. But this is exactly what Wilde's model implies. Clearly, in the large number case, the free-rider problem undermines his analysis.

In Wilde's single person model, the individual works at even a 100 percent tax rate because *his* work efforts determine the availability of Z (the alleged public good). Of course, it is plausible for Robinson Crusoe to work harder, in order to supply more national defense. Crusoe is decisive in determining how much defense his island has. Wilde's model illustrates this point. However, in the large number case, individuals have little incentive to work more because their individual effort exerts virtually no impact on the supply of the public good.

Similarly, the absurdity of a monotonic Laffer curve would not emerge from a properly formulated model with a large population. In Wilde's model, higher tax rates induce the individual to work more since the high rate increases *his* ability to transform leisure into Z. This factor, along with a negative income effect (in utility terms) explains, why the higher tax rates lead to *increases* in both hours worked and tax revenues over a wide range of very high tax rates in Wilde's model. However, in a large population model, the quantity of public goods would be largely independent of the *individual's* work effort. Thus, the source which generates the additional work effort leading to the monotonic Laffer curve would be absent.

In conclusion, Wilde's paper is inapplicable to an economy with a large number of individuals. It neither adds to nor detracts from the central point of our original article —the internal inconsistency involved in applying the traditional work-leisure analysis for the individual to the aggregate economy.

REFERENCES

Browning, Edgar K. and Browning, Jacquelene M., *Public Finance and the Price System*, 2d ed., New York: Macmillan, 1983.

Gwartney, James and Stroup, Richard, "Labor Supply and Tax Rates: A Correction of the Record," *American Economic Review*, June 1983, *73*, 446–51.

Wilde, James A. "Labor Supply and Tax Rates: Comment," *American Economic Review*, December 1984, *74*, 1103–07.

An Empirical Test of the Infant Industry Argument: Comment

By ROBERT E. B. LUCAS*

Anne Krueger and Baran Tunçer (1982) present some interesting results on rates of effective protection and of productivity growth for various sectors of the Turkish economy. From these measures, the authors "develop an empirical test for the validity of the infant industry argument" (p. 1149). Being perhaps the first attempt to test directly the commonly cited infant industry grounds for protection, the paper by Krueger and Tunçer is especially valuable. Clearly, the approach is readily applicable in other contexts for further testing but, before so doing, it seems worth pursuing certain implications of that test a little more precisely.

The test proposed by Krueger-Tuncer for the validity of the infant industry argument is founded on the premise that "a necessary (but not sufficient) condition is that costs in (temporarily) assisted or protected industries should have fallen over time more rapidly than costs in nonprotected or less-protected industries" (p. 1144). "The test is simple and straightforward: input per unit of output must fall more rapidly in more protected industries if there is to be any rationale for infant industry protection. In the Turkish case, there was no such tendency over the period covered" (p. 1149). But is this a sufficient test of the necessary conditions?

The case for infant industry assistance is derived either from an externality associated with learning-by-others-doing or learning-by-own-doing combined with a capital market failure. For the instance of within industry learning, this is commonly modelled by inserting the integral of past output as an argument in the production function, which may be written for industry i at time t:

$$Q_{it}=Q^{i}\left(L_{it},\int_{\tau=0}^{t}Q_{i\tau},t\right), \tag{1}$$

where Q is output and L is a vector of primary and intermediate inputs. Abstracting from technical progress other than that embodied in cumulative output, and imposing a suitably convenient functional form, (1) may be rewritten

$$q_{it}=q^{i}(\textstyle\int Q_{i\tau}), \tag{2}$$

where q is output per unit of some total input index. Now add a supply function for commodity i by domestic firms:

$$Q_{it}=s^{i}(e_{it},z_{it}), \tag{3}$$

where e is the effective rate of protection, and z is a vector of other elements shifting supply (world prices, commercial policy in other sectors, prices of nontradeables, etc.).[1] Thus, the rate of change in factor productivity is

$$\hat{q}_{it}=\left(q_{1}^{i}/q_{it}\right)\cdot s^{i}(e_{it},z_{it}), \tag{4}$$

where q_1^i is the first derivative of (2).

Obviously, if q_1^i and s_1^i, (the first derivative of s^i with respect to e), are positive, then the higher the effective protection on i the greater should be the rate of factor productivity increase.[2] This is the foundation of the Krueger-Tuncer test. But it does not necessarily follow that sectors with higher

*Center for Asian Development Studies, Boston University, Boston, MA 02215. I am particularly grateful to Anne Krueger for correspondence on this topic, and to my colleagues P. Cory, R. Ellis, S. Hunt, D. Schydlowsky, and P. Tandon for comments.

[1] More generally, if factors are imperfectly mobile between sectors, and in particular if capital (either physical or human) is nonmalleable, lagged values of e_i ought also to be arguments of (3).

[2] Note that this statement remains correct even if nonlearning technical progress is not assumed away.

rates of effective protection should have more rapid productivity increases. Thus:

$$e_{it} > e_{jt} \nRightarrow \dot{q}_{it} > \dot{q}_{jt} \qquad i \neq j. \tag{5}$$

A sufficient condition for (5) to hold would be common functions for q_1^i and s^i, across all sectors, which is quite unlikely to be true. The effect of cumulative output on learning will vary from sector to sector; the supply responsiveness of output to protection will be sector specific; and the levels of z as well as its effects on supply will also differ between sectors. The Krueger-Tuncer test is not sufficient to reject the necessary condition for infant industry learning in this model, that $q_1^i > 0$.

Indeed, if the model is extended to include learning-by-others-doing, even a positive value for q_1^i is not strictly required. Positive cross-sector learning could, in principle, suffice as a basis (though not sufficient condition) instead.

But even granted circumstances when $q_1^i > 0$ is critical, provided some learning is engendered by protection, it is not necessarily true such learning must be more rapid "in more protected industries if there is to be any rationale for infant industry protection" (p. 1149). That rationality depends upon intertemporal costs and benefits. The benefits in slow learning sectors are delayed, but assistance may nonetheless be warranted, depending upon the sectoral specific costs of any given level of protection and the rate of time preference.

In the end, only a careful weighing of intertemporal, social costs and benefits can discern whether infant industry protection might be justified. In any such evaluation of dynamic comparative advantage, a critical set of parameters must be the extent of learning induced by protection within various sectors. But such parameters cannot be appropriately identified and estimated, nor even their average positiveness reliably tested, from cross-sectoral data alone: each individual within-sector response may be positive yet the cross section appear negative. Meanwhile, perhaps the onus should fall on demonstrating rather than assuming sufficient grounds: the world is certainly littered with geriatric infants.

REFERENCE

Krueger, Anne O. and Tuncer, Baran, "An Empirical Test of the Infant Industry Argument," *American Economic Review*, December 1982, *72*, 1142–52.

An Empirical Test of the Infant Industry Argument: Reply

By Anne O. Krueger and Baran Tuncer*

Robert Lucas raises two points with respect to our earlier article (1982a): 1) because the time path of learning with respect to cumulative output might differ across industries, he asserts that it might be warranted to have higher protection in a slower-learning industry (and hence that our test might fail); and 2) because the externalities of an infant industry might accrue to activities other than that industry, the direct test of factor productivity growth may be unfair. The two points are fundamentally different, so we deal with each in turn.

With respect to the first point, Lucas posits that the rate of total factor increase in each industry might be a function of cumulative output, but that the shapes of the functions might differ. He fails, however, to note that the costs of protection in each period of time are the area of the triangle equal to one-half times the rate of effective protection times output. Hence, the cumulative costs of protection are one-half times the integral of the effective protection rate times output.

Note three things: 1) the higher the costs, the higher must be the net benefits. Hence, the higher effective protection, the more costs must fall, although not necessarily year by year. This is why we chose a longer time period over which to conduct our test. 2) If a slow learner warranted more protection, it could only be in the case where learning would be faster at a later date, since otherwise it would pay to pour more resources into the faster learner and fewer into the slower; 3) Lucas's own judgment, that "only a careful weighing of intertemporal, social costs and benefits can discern whether infant industry protection might be justified" (p. 1111) agrees with ours once it is recognized that the costs are the excess cost of producing behind a wall of protection, and the only conceivable benefits are the cost reductions over time; and 4) Lucas's comment about cross-section observations being an invalid measure quite agrees with our own: that is why we took effective rates of protection (which were fairly constant over a long period of time) and associated them with rates of total factor productivity growth over periods of a decade or more. Moreover, as inspection of our results by subperiod (given in our article, 1982b) demonstrates, there was certainly no tendency for *TFP* growth rates to accelerate or decelerate in any systematically different way across industries.

We, therefore, interpret Lucas to be saying that our time period was insufficient to warrant our results. When 80 percent or more of new investments are directed to industries whose costs exceed world costs by 50 percent or more for a decade and a half, one wonders what (low) rate of time preference could possibly warrant those excess costs. As our computations indicated, even for the industry with the most rapid productivity growth, 24 years would be required at that rate in order to break even. We did not even attempt to calculate how much longer productivity would have to grow at that rate before costs of earlier protection were recovered, largely because it was our judgment that the real rate of return (and therefore opportunity cost of capital) in Turkey was well in excess of 5 percent—a level at which costs could never be recovered.

As to Lucas's second point, he is quite correct. If externalities accrue outside the protected industry, protection might be warranted. That is why we computed economy-wide rates of total factor productivity growth, because surely those externalities should show up somewhere.

*The World Bank, 1818 H Street NW, Washington, D.C. 20433.

REFERENCES

Krueger, Anne O. and Tuncer, Baran, (1982a) "An Empirical Test of the Infant Industry Argument," *American Economic Review*, December 1982, *72*, 1142–52.

______ and ______, (1982b) "Growth of Factor Productivity in Turkish Manufacturing Industries," *Journal of Development Economics*, December 1982, *11*, 307–25.

Lucas, Robert E. B., "An Empirical Test of the Infant Industry Argument: Comment," *American Economic Review*, December 1984, *74*, 1110–11.

The Use of Inputs by the Federal Reserve System: Comment

By JOHN H. BOYD*

In a recent issue of this *Review* (1983), William Shughart and Robert Tollison (S-T) hypothesize that the Fed pursues a bureaucratic objective which results in an "inflationary bias" in monetary policy. They test this hypothesis empirically and conclude that it is supported by the data. This comment, however, shows that their statistical procedures are flawed and when corrected provide little support for the inflationary bias theory. The data actually suggest a different, but not very interesting or surprising, interpretation.

I. Regression Results

Shughart-Tollison hypothesize that monetary policy is carried out to achieve a bureaucratic objective which includes as one argument the value of amenities (wages, salaries, perquisites, etc.) to Federal Reserve employees. Such self-dealing is constrained by the requirement that the Fed turn over its revenue, after expenses, to the Treasury. The Fed therefore faces a tradeoff: it can allocate resources to its own employees in the form of amenities, or to the public in the form of reduced taxes. This budget constraint can be relaxed, however, if the Fed makes open market purchases that add to its security portfolio and increase interest income. Therefore, self-dealing may result in an inflationary bias, causing the Federal Reserve System to provide more reserves to the banking system than dictated by public policy objectives.

Although this theory is not directly testable, one of its predictions is: if the theory is correct, the monetary base will be positively associated with the total value of amenities to Fed employees.

Using Federal Reserve System employment as a proxy for Fed amenities, S-T attempt to test this prediction. They estimate an equation like (1) below, with the nominal monetary base ($BASE$) as the dependent variable and Fed employment ($SYSEMP$) as one of several independent variables. As predicted, the coefficient of $SYSEMP$ is positive and is significantly different from zero at a high confidence level (t-values are shown in parentheses).

$$(1)\quad \begin{aligned} BASE = -1.00 &+ \underset{(2.62)}{.13}\, SYSEMP + \underset{(3.18)}{.34}\, CD \\ &- \underset{(.75)}{.01}\, INT_{-1} + \underset{(2.00)}{.31}\, RGNP \\ &+ \underset{(6.18)}{.06}\, TREND - \underset{(1.83)}{.0002}\, TREND^2 \end{aligned}$$

$$n = 66;\quad R^2 = .93;\quad \hat{\rho} = .77;$$

$$D\text{-}W = .89;\quad Q(24) = 126$$

where $BASE$ = the nominal monetary base; $SYSEMP$ = the number of Federal Reserve System employees; CD = the currency/deposit ratio; INT = the interest rate on 4–6-month prime commercial paper; $RGNP$ = real gross national product; $TREND$ = a linear time trend. All variables except INT and $TREND$ are natural logs. The sample period, 1916–81, is estimated with generalized least squares.

Actually, (1) is not S-T's estimate, but is my own attempt to replicate their results. The two estimates are not identical, but almost so. All regression coefficients and test statistics are the same to two digits except for R^2 where my estimate is .93, theirs .95. This is close enough for government work.

*Federal Reserve Bank of Minneapolis and University of Minnesota, Minneapolis, MN 55455. I was ably assisted in the preparation of this comment by Danny Quah, who did the statistical work and patiently instructed me in modern econometrics. Bruce Smith also provided helpful comments. The views expressed herein are my own and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

I have two observations about this regression. First, it is not particularly robust to specification changes. S-T are very careful to interpret (1) as a supply function representing Fed policy actions. If this is the correct interpretation, I find it hard to believe that, over this time period, the Fed responded only to changes in real *GNP* and not to changes in the price level. Yet, if the price level is included on the right-hand side, either as an independent term or by replacing real with nominal *GNP*, the coefficient of *SYSEMP* becomes insignificantly different from zero. This is bothersome since that coefficient is the key parameter estimate.[1]

Second, S-T do not report Durbin-Watson (*D-W*) or *Q* statistics. They should, though, because even following their procedure to correct for serial correlation, the transformed errors exhibit strong positive autocorrelation. In my estimate, the *D-W* statistic is .89 and the *Q* statistic an alarming 126.[2] This leads one to be dubious of the reported *t*-statistics.

II. Tests of Causality

Shughart-Tollison do not explicitly model the Fed's decision process and, therefore, can make no tight prediction about the timing of changes in *SYSEMP* and *BASE*. Even so, causality (in the Granger sense) is important, and if their hypothesis is to receive any compelling support, it is required that *SYSEMP* cause *BASE* and not vice versa. That is so because there is another good reason—having nothing to do with their theory—to expect these two variables to be positively correlated. This other explanation, however, would predict that *BASE* causes *SYSEMP*. I will return to it momentarily.

Shughart-Tollison are aware of the problem and realize that, in itself, positive contemporaneous correlation between *SYSEMP* and *BASE* (if, indeed, they have demonstrated that) provides little support for their hypothesis. Therefore, they carry out some Sims tests of Weiner-Granger causality (Christopher Sims, 1972). These indicate that causality is from *SYSEMP* to *BASE* and, moreover, that the hypothesis of no reverse causality (from *BASE* to *SYSEMP*) cannot be rejected. From this, they conclude that

> ...the appropriate regression specification is one in which the monetary base is the dependent variable. That is, employment belongs in the money supply equation; the base does not belong in an input demand function. This result provides further support for the notion that bureaucratic incentives drive the money supply process and contradict the alternative view that increases in the base result in more employee hiring for the purpose of handling larger numbers of accounting entries. [p. 302]

In these tests S-T use leading and lagging values of the two variables *SYSEMP* and *BASE* only and in support of this procedure cite Sims. But, as John Geweke (1978) and others have pointed out, a two-variable test may not be appropriate. If there are additional variables in the information set, the appropriate test procedure is to include them all. Yet, it is obvious from the specification of (1) that S-T believe such other variables exist. Otherwise, why include them in the multivariate regression? Therefore, I have redone the causality tests in a way that is consistent with their own multivariate specification. Because of the difficulty in carrying out multivariate Sims tests, however, I have instead used the Granger method. The two procedures are asymptotically equivalent.

Shown in Table 1 are the results of tests in which *BASE* is regressed on lagged values of itself and of all other variables in equation (1). Table 2 is identical except that it reports tests with *SYSEMP* dependent. In both cases, two lag lengths are arbitrarily chosen, three and six years. In addition to estimates using a nonlinear time trend as S-T do, there are estimates with no trend variables and

[1] For example, when real *GNP* is replaced with nominal *GNP* in (1), all else unchanged, the coefficient of *SYSEMP* is .040 and its *t*-value is .746. S-T report different results when price is included (fn. 12, p. 299), but do not indicate what structural form they used.

[2] The *Q* statistic (distributed *Chi*-squared) provides on overall test on the autocorrelation of estimated residuals at all lags. In this case, the null hypothesis, that the true residuals are white noise, is rejected at the 99.99 percent confidence level.

TABLE 1—COEFFICIENTS OF *SYSEMP* IN REGRESSIONS WITH *BASE* THE DEPENDENT VARIABLE[a]

	Lag Length						
	−1	−2	−3	−4	−5	−6	F
	−.05	.22	−.41				5.62[b]
TT	.03	.23	−.44				3.89[b]
PW	.02	.14	−.08				.54
	−.03	.05	−.22	−.09	.11	−.15	3.45[b]
TT	.04	−.03	−.06	−.20	.17	−.06	1.06
PW	−.04	.05	.01	−.18	.09	.05	.89

[a]*TT* indicates that *TREND* and $TREND^2$ were included as explanatory variables; *PW*, that the data were prewhitened. The F test is against the null hypothesis that all (3 or 6) coefficients of *SYSEMP* or *BASE* are zero.
[b]Indicates rejection at the 90 percent confidence level or higher.

TABLE 2—COEFFICIENTS OF *BASE* IN REGRESSIONS WITH *SYSEMP* THE DEPENDENT VARIABLE[a]

	Lag Length						
	−1	−2	−3	−4	−5	−6	F
	.00	.17	−.00				2.31[b]
TT	.03	.10	.08				3.17[b]
PW	−.10	−.12	.46				3.66[b]
	−.19	.38	.60	−.51	.28	−.20	3.45[b]
TT	.30	.28	.57	−.32	−.06	−.09	6.06[b]
PW	.03	.28	.60	.19	−.11	−.41	2.47[b]

[a,b]See Table 1.

also estimates in which the data are prewhitened using a filter suggested by Sims. For brevity, I report only regression coefficients for the variable where causation is at issue. And for comparability, the same effective sample size is used with both lag lengths.

Table 1 indicates that lagged values of *SYSEMP* help predict *BASE* in three of the six specifications. In the three others, however, one cannot reject the hypothesis of no causality. The results seem quite sensitive to specification and are best described as "ambiguous." That is, with these sample data it's just not clear if *SYSEMP* does...or does not...cause *BASE*.[3] Table 2 is another story. In all six specifications, the null hypothesis, that *BASE* does not cause *SYSEMP*, is rejected at the 90 percent confidence level or higher. In summary, my results are quite different from those reported by S-T.[4]

As mentioned earlier, causality running from *BASE* to *SYSEMP* is consistent with a demand relationship. The idea is that, as economic activity and the demand for money balance increase, so will the demand for *payment services* (coin, currency, check clearing, etc.) provided by the Federal Reserve. Producing payment services requires labor input; in fact, it is quite labor intensive, accounting for about 80 percent of total Fed employment. These employees are primarily production-type workers who operate check sorting machines, physically handle coin and currency, etc. As the demand for money

[3]It is also interesting that prewhitening seems to have a significant effect with these data. (According to results reported by Geweke, p. 181, that should not be the case.)

[4]These results suggest another problem with equation (1), in addition to those already mentioned; that is, the dependent variable causes one of the explanatory variables.

increases, demand for these services increases also and the Fed responds by hiring more workers. It is plausible that it responds with some lag, and this is what the data seem to suggest.[5]

III. Conclusion

It is possible that there does exist an "inflationary bias" in Fed policy. However, the statistical results S-T have presented are simply not compelling. In particular, an important part of their study is to show that *SYSEMP* causes *BASE*, a result that would seem implausible absent their theory of bureaucratic behavior. My tests cast doubt on this causal relationship. But even if they are (were) right, it turns out that there are several *other* variables which also appear to be strongly affected by *SYSEMP*—despite similar a priori implausibility.

[5]Until 1981, there was no charge for these payment services, other than the fixed cost to banks of being Federal Reserve members. In effect, the Fed passively provided whatever quantity was demanded by the banking sector. Naturally, the relationship between the demand for payment services and the Fed's derived demand for labor inputs would be blurred by changes in technology and by changes in the menu of services available, and over the sample period there were many of each.

For example, in one or more of the specifications discussed earlier, the variable *SYSEMP* causes the currency/deposit ratio as well as interest rates, at the 95 percent confidence level or higher. Now, on the basis of their tests, S-T have claimed that "...every time the Fed hired one more worker, the stock of high-powered money rose by about $362,000" (p. 300). On the basis of statistical procedures at least as defensible and arguably superior, I could claim that "Every time the Fed hired one more worker, the currency/deposit ratio declined by about .00026 percent and interest rates declined by about .013 basis points." I doubt that anyone would take my claim seriously, nor, frankly, would I expect them to.

REFERENCES

Geweke, John, "Testing the Exogeneity Specification in the Complete Dynamic Simultaneous Equation Model," *Journal of Econometrics*, April 1978, *7*, 163–85.

Shughart, William F., II and Tollison, Robert D., "Preliminary Evidence on the Use of Inputs by the Federal Reserve System," *American Economic Review*, June 1983, *73*, 291–304.

Sims, Christopher A., "Money, Income, and Causality," *American Economic Review*, September 1972, *62*, 540–52.

The Use of Inputs by the Federal Reserve System: Comment

By JOHN S. STRONG*

In a recent article in this *Review* (1983), William Shughart and Robert Tollison (S-T) argue that an important motivation for expansion of the money supply is to finance the growth of the Fed's bureaucracy: specifically, that the level of Fed employment is a determinant of the nominal monetary base. Using a public choice framework, they present a money supply equation which yields a positive and statistically significant relationship between the monetary base and the total employment of the Federal Reserve System. While this conclusion is certainly interesting, the analysis suffers from a number of important conceptual, empirical, and methodological problems.

Shughart-Tollison point out that the revenue side of the Federal Reserve's budget is closely linked to its monetary policy operations. Through open market purchases of securities, the Fed generates a stream of interest payments. Because monitoring of this income stream is difficult and costly, Congress has elected to impose the constraint that "excess revenues" be returned to the Treasury. As a result, the Fed pays its current operating expenses out of this current interest income and then turns over the remainder to the Treasury. Because the central bank cannot retain the profits it generates, S-T argue that the Fed will exhibit a form of expense-preference behavior in purchasing more amenities than are justified by cost-minimizing money production—in particular, by expanding Fed employment. In addition, since expansionary monetary policy increases opportunities for this profit/amenities tradeoff, Fed policies are likely to have a built-in inflationary bias.

To test this hypothesis, S-T develop a regression model that estimates the monetary base as a function of Fed employment, real *GNP*, a number of financial variables, and linear and nonlinear time trend variables. Using annual time-series data from 1915 to 1981, they find that the employment variable is a significant determinant of the money supply, with their model explaining about 95 percent of the variation in the monetary base. To check that employment causes monetary expansion and not the other way around, S-T employ the test of causality developed by C. W. J. Granger (1969) and Christopher Sims (1972). This test estimates two-sided distributed lag regressions and compares results. That is, the current monetary base is regressed on past, present, and future Fed employment figures; then, for comparison, current Fed employment is regressed on past, present, and future values of the monetary base. Fed employment in the previous period was found to be a significant determinant of the current monetary base; no such relation was found for past values of the monetary base with regard to current Fed employment. In addition, the future value of the monetary base was statistically related to current Fed employment. On these grounds, S-T conclude that central bank employment does belong in the money supply equation; that is, that increases in Fed employment are a cause of money supply growth.

However, both the original model and the test of causality have serious problems. While the original model regressed the monetary base on Fed employment, a regression of Fed employment on the monetary base yields equally good results, thus casting doubt on the direction of causality. Moreover, most of the S-T model's explanatory power comes from the inclusion of the time trend variables (as is the case in many time-series analyses). In fact, a simple regression of the monetary base on the time trend and (time trend)^2 variables generates values of corrected R^2 on the order of .99.

The results claimed from the use of the Granger-Sims test of causality are also ques-

*Harvard University, Cambridge, MA 02138.

tionable. The significance levels are very low in both sets of regressions, so that it would not be difficult to interpret the results as saying that there is little or no connection between system employment and the monetary base in either direction. In addition, careful interpretation of the Granger-Sims results yields a different interpretation than that given by S-T. In their original model, S-T regressed the current monetary base on current system employment, and found a statistically significant relation. However, when this relationship is evaluated in the Granger-Sims framework, there is no evidence of a contemporaneous relationship (the t-statistics are 0.14 and -0.04). If past levels of Fed employment are determinants of the current monetary base, then S-T's original model is misspecified; lagged values of employment should be included as independent variables. When this lagged regression is performed, however, the posited relationship is no longer evident.

The central issue in the S-T paper is the relation between correlation and causation. The money supply function summarizes in a sophisticated way the historical correlations of economic time-series. While correlation per se may be enough for forecasting, the use of such models as explanators of economic phenomena is much more open to question. As James Tobin (1970) has pointed out, even if variable X leads variable Y (i.e., X is correlated with future Y) no causality is implied. Moreover, Sims (1980, 1982) himself shows that even if X "Granger-causes" Y (i.e., helps to predict Y), it provides no evidence of a causal relation between the two variables. Rather, X may be correlated with other variables or unobservables which themselves cause Y.

A theoretical note of caution should also be mentioned. S-T argue that the employment effect serves as an additional determinant of the stock of high-powered money. If this is true, then it is appropriate to consider the problems that result from attempting to incorporate the employment variable into the monetary sectors of large-scale macro models. The dangers of one-equation-at-a-time estimation of components of macroeconomic systems (of which the money supply is clearly relevant) have been pointed out by Sims (1980) and Zvi Griliches (1968). The behavioral implications of the restrictions on all sectoral equations may be much less reasonable than the restrictions on any one equation taken by itself. In this context, single-equation specification is likely to lead to overidentified models and spurious behavioral interpretations. These identifying restrictions serve to distort the structural relationships and may thus be of no value in conducting formal statistical tests of macroeconomic theories (such as the S-T model). In this context, empirical research should be careful to avoid testing macroeconomic theories in a single- or few-equation framework.

The expense-preference description of Fed employment growth is open to serious question. An alternative view, and in my mind a more plausible one, is that the secular growth in Fed employment is the result of a continual expansion of Federal Reserve functions since 1915. An analysis of the growth pattern of employment indicates that most of the increases have been concentrated in four periods: the initial years 1915–20; the early 1930's, related to the Great Depression; World War II; and the early 1970's, which involved an expansion of bank regulatory and supervisory functions. While it is difficult to separate discrete and continuous increases in employment, it is important to focus primary attention on the scale and scope of Federal Reserve functions as sources of employment growth, rather than applying expense-preference theory to aggregate employment totals.

The theoretical frameworks of public choice economics and of principal-agent relations promise to make useful contributions to our knowledge about economic policymaking. However, empirical tests of these approaches should explicitly recognize the behavioral restrictions that these models place on the rest of the macroeconomic system. In addition, the highly correlated nature and secular growth trends that characterize most economic time-series raise serious questions about the ability of single equation structural models to validate behavioral hypotheses.

The Use of Inputs by the Federal Reserve System: Reply

By WILLIAM F. SHUGHART II AND ROBERT D. TOLLISON*

In our 1983 paper we offered evidence that the law of demand operates inside the Federal Reserve. In particular, the Fed faces a requirement that it return all revenues in excess of operating expenses to the Treasury, and this constraint lowers the price of amenities in terms of foregone profits. The monetary authority accordingly buys more of the wage and nonwage perquisites of office than otherwise. Because we treated amenities as a monotone transformation of Federal Reserve System employment, our theory suggested that the Fed would pad its operating expenses by increasing the number of employees on its payroll. Moreover, given that expansionary open market operations raise the interest income earned by the Fed on its securities portfolio, bureaucratic incentives would impart an inflationary bias to monetary policy.

In subsequent tests of the theory, we found a positive and significant *ceteris paribus* relationship between changes in the monetary base and the size of the Fed. This result suggested that one motivation for expansions in the money supply is to finance the growth in the Fed's bureaucracy. We also found evidence that employment causes money in the sense of Christopher Sims (1972), but not the reverse, and that the growth in Fed employment over time does not appear to have been due to the fact that more people are required to manage larger money stocks.

In their comments, John Boyd and John Strong suggest that there are methodological and empirical problems with our paper. Both comments focus primarily on the causality tests, but each raises other issues designed to cast doubt on the strength of our results. In what follows, we discuss the main points raised by our critics.

*Clemson University, Clemson, SC 29631, and Center for the Study of Public Choice, George Mason University, Fairfax, VA 22030, respectively.

I. Causality

To test whether or not it was appropriate to include Fed employment as an independent variable in our money supply equation, we employed the test of causality developed by Sims in extending the work of C. W. J. Granger (1969). Sims' test for unidirectional causality involves estimating paired two-sided distributed lag regressions, then comparing the relative sizes of the coefficients obtained on past and future values of the independent variable as well as determining whether, taken as a group, the coefficients on the future values are significantly different from zero.

Our implementation of Sims' technique was as follows. We first detrended the money supply and employment series. Then for each possible dependent variable we specified one regression containing contemporaneous and two past values of the other variable, and a second regression augmented by two future values. When we regressed the monetary base on past and future Fed employment, the relevant F-statistic implied that the future employment variables did not add significantly to explaining the variation in the monetary base, and the coefficients on past and future employment variables were "small." In contrast, the coefficients obtained when future monetary base variables were added to the regression of employment on money were significantly different from zero as a group. Moreover, the coefficients on future money were "large" compared with the coefficients on the lagged variables. Thus we concluded that there was evidence of unidirectional causality running from employment to money, that is, that employment belongs in the money supply equation; the base does not belong in an input demand function.

Strong argues that the central issue in our paper is the distinction between correlation and causation. He cites three pieces of information that suggest to him that we have not

established the direction of causality. These are that most of the explanatory power of our money supply equation comes from the inclusion of linear and nonlinear trend variables, that significance levels are very low in both sets of causality regressions, and that the *t*-statistics in the causality tests imply no contemporaneous relationship between money and employment.

Strong's first point is readily dealt with. When we estimated what we thought to be an appropriately specified money supply equation, Federal Reserve System employment entered with a positive and statistically significant coefficient. In the absence of a theory, the fact that when employment and other variables are dropped the time trends still explain a large proportion of the variation in the monetary base is irrelevant. What matters is that there is evidence of a *ceteris paribus* relationship between the money stock and Fed employment.

As to Strong's other points, Sims stresses that one should not place great weight on significance levels. He states in part that "it is a truism too often ignored that coefficients which are 'large' from the economic point of view should not be casually set to zero no matter how statistically 'insignificant' they are," and that "small coefficients on future values of the independent variable may sometimes be safely ignored even when they are statistically significant" (pp. 545–46). The crucial elements of Sims' test are the relative sizes of lagged and leading coefficients and the significance of the coefficients on future values, taken as a group. On this basis, our results establish the direction of causality insofar as the technology of the test allows one to do so. Strong has simply misread Sims.

We find similar fault with the criticisms raised by Boyd regarding causality. Boyd suggests that we erred by omitting the other exogenous variables in our money supply equation—*GNP*, the currency-deposit ratio, price level, and so forth—when estimating the two-sided distributed lag regressions. Our interpretation of Sims on this issue is that the inclusion of additional independent variables is likely to make a unidirectional causal structure appear bidirectional (Sims, p. 543). Moreover, in attempting to "correct" our results for this omission, Boyd discards Sims' test in favor of the more easily implemented method of Granger. He finds bidirectional causality.

II. The Demand-Side Alternative and Other Issues

Both Boyd and Strong offer alternative explanations for our results which focus on the demand for inputs by the Fed. Boyd suggests that as the demand for money increases, so will the demand for the services provided by the monetary authority—coin, currency, check clearing, and so on. Moreover, these services are quite labor intensive according to Boyd. Similarly, Strong believes that "the secular growth in Fed employment is the result of continual expansion of Federal Reserve functions since 1915" (p. 1119). He argues that since most of the increases in Fed employment have been concentrated in four historical periods, something like increased bank regulatory and supervisory duties may lie behind the observed expansion of the central bank.

Neither critic provides any empirical support for his alternative hypothesis, nor discusses our evidence on this point (pp. 300–01). This is particularly disturbing in regard to Boyd's demand-side explanation since it stands in direct contradiction to the standard assumption that there are substantial scale economies in money production. On the other hand, Strong's suggestion may have some merit, and it would be interesting to have a test of his theory. Our results do suggest that the Fed is "different," and we would expect this difference to carry over to the monetary authority's other functions. For example, our theory would imply that Fed employment devoted to enforcement of the Equal Credit Opportunity Act is larger than if the law were enforced by an independent agency subject to the normal budgetary process. We think this is a worthy subject for future research.

Boyd further claims that our money supply regression results are not particularly robust

to specification changes.[1] Specifically, he finds it hard to believe that we did not include the price level as an independent variable. We did report results from including the implicit *GNP* price deflator on the right-hand side of the regression in our footnote 12 (p. 299), however. In addition, we noted that our results were not changed qualitatively by using relative Fed employment as an explanatory variable, by putting all variables in percentage change terms, or by including the velocity of money as an independent variable.

A final point raised by Strong is that we should have incorporated our employment effect into the monetary sector of a large-scale macroeconomic model. We frankly have no interest in this problem. Our bias is clearly toward a micro explanation of Fed behavior, and our main point was simply to provide evidence that the many policy "failures" recounted by the Fed's critics are the result of understandable behavior of self-interested maximizers operating under given constraints.

In sum, we find nothing in the comments of Boyd and Strong that would lead us to alter our original paper.[2] We reemphasize, however, that we have not argued that the tail wags the dog, that is, that the employment effect drives the money supply process. We suggest, rather, that employment is one of many margins on which the Fed operates, and that working along that margin may tip monetary policy toward expansionism.

[1]On a minor point, Boyd chides us for not reporting Durbin-Watson (D-W) statistics for our estimated money supply equations. The null hypothesis for the D-W d is, of course, the absence of first-order autoregression. The statistic is therefore meaningless when a correction for serial correlation has already been applied.

[2]It is simply silly for Boyd to conclude by saying that his "arguably superior" empirics suggest a causal relationship running from Fed employment to either the currency-deposit ratio or to interest rates. We have a theory; he does not.

REFERENCES

Boyd, John H., "The Use of Inputs by the Federal Reserve System: Comment," *American Economic Review*, December 1984, *74*, 1114–17.

Granger, C. W. J., "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, July 1969, *37*, 424–38.

Shughart, William F. II and Tollison, Robert D., "Preliminary Evidence on the Use of Inputs by the Federal Reserve System," *American Economic Review*, June 1983, *73*, 291–304.

Sims, Christopher A., "Money, Income, and Causality," *American Economic Review*, September 1972, *62*, 540–52.

Strong, John S., "The Use of Inputs by the Federal Reserve System: Comment," *American Economic Review*, December 1984, *74*, 1118–20.

Illusory Wage Differentials: Comment

By KEITH HYLTON*

A recent article in this *Review* (1979) by Edward Lazear contends that black-white wage differentials have not narrowed as much as recent data may indicate. He suggests that the finding of a narrowing differential results in part from a mismeasurement of the real wage. Using a new measure, which supposedly takes account of differences in on-the-job training (*OJT*), Lazear finds that whites enjoy a higher relative wage advantage than observed (1.47 as opposed to 1.37). This comment focuses on the new wage measure. When estimated with data from different time periods in the *National Longitudinal Survey* (*NLS*) sample, it is found to be unstable.

The proposed measure is the observed wage plus some approximation of the amount by which wage is increased by experience. Labeling the corrected wage in the tth year of experience w_t^c and the observed wage w_t, the measure is expressed

$$(1)\quad w_t^c = w_t + \int_0^{N-t} (dw/dt)\, e^{-rv}\, dv,$$

where N represents the total number of years in the labor force and r, constant rate of interest. The latter term in (1) may be considered as the unobserved *OJT* component of the wage. Lazear contends that as wage differentials have narrowed, the gap between average *OJT* components has widened, thus leaving earnings differentials, in a present value sense, unaltered.

Letting w_t^{Bc} and w_t^{Wc} represent corrected black and white wages, the corrected relative wage of whites is[1]

$$(2)\quad \frac{w_t^{Wc}}{w_t^{Bc}} = \frac{w_t^W + \int_0^{N-t} (dw^W/dt)\, e^{-rv}\, dv}{w_t^B + \int_0^{N-t} (dw^B/dt)\, e^{-rv}\, dv}.$$

One can readily show that (2) varies as

$$(3)\quad w_t^W / w_t^B \lesseqgtr (dw^W/dt)/(dw^B/dt).$$

I. Estimation of Wage Growth Model

To obtain an estimate of dw/dt, Lazear uses a wage growth equation of the following

*Graduate student, Economics Department, Massachusetts Institute of Technology, Cambridge, MA 02139. I thank John Bound, among others, for helpful comments.

[1]It is interesting to note that under certain conditions (2) serves as an approximation of relative human wealth. Given that white and black workers work similar hours, the relative human wealth of a white worker can be expressed as the following ratio:

$$\int_0^N w^W(t)\, e^{-rt}\, dt \Big/ \int_0^N w^B(t)\, e^{-rt}\, dt.$$

Expanding $w(t)$ around zero,

$$\int_0^N w(t) e^{-rt} \cong \int_0^N [w(0) + w'(0)t]\, e^{-rt}\, dt$$

$$= \frac{w(0)(1-e^{-rN})}{r}$$

$$+ \frac{w'(0)[1-e^{-rN}(rN-1)]}{r^2}$$

$$\cong \frac{1}{r}\left[w(0) + \frac{w'(0)}{r}\right]$$

for large N. The human wealth ratio is then $[w^W(0) + w^{W\prime}(0)/r]/[w^B(0) + w^{B\prime}(0)/r]$, which for large values of N is roughly the same as (2).

TABLE 1—ESTIMATES OF THE WAGE GROWTH MODEL[a]

Variable	Lazear's Results 1966–69	New Estimates 1966–69	1967–70	1968–71	1966–71
Intercept	.586	.4744	.5126	.1490	.1517
	(4.69)	(3.82)	(2.92)	(.78)	(.56)
A_t	−.02097	−.01694	−.00861	−.00169	−.01536
	(4.39)	(3.54)	(1.57)	(.36)	(2.27)
E_t	−.01445	−.00985	−.01345	−.00714	−.01758
	(2.80)	(1.95)	(2.37)	(1.52)	(2.47)
ΔH	.00178	−.00120	−.00440	−.00419	−.00327
	(2.66)	(2.02)	(5.22)	(5.16)	(3.34)
ΔS	.06288	−.05657	−.00774	.09825	.08283
	(1.91)	(1.78)	(.14)	(1.32)	(1.34)
ΔST	−.14681	−.17918	−.19850	−.16602	−.25254
	(6.75)	(8.50)	(6.71)	(6.03)	(7.27)
M_t	−.09688	−.04897	−.06011	−.08358	−.06401
	(3.44)	(1.67)	(1.69)	(2.80)	(1.65)
M_τ	.01287	.01727	.05162	.00756	.00629
	(.52)	(.68)	(1.69)	(1.37)	(.91)
ΔE	.09286	.08432	.01215	.04762	.13428
	(2.50)	(2.92)	(.28)	(1.19)	(2.94)
U_τ	.50531	.47046	−.39150	.14733	.34138
	(4.97)	(6.34)	(3.00)	(.83)	(1.73)
D	−.08262	−.01680	.10091	.33314	.03620
	(1.10)	(.23)	(.80)	(2.05)	(.18)
$(\Delta S)(\Delta E)$	−.00615	−.00896	−.00607	−.01594	−.00959
	(.46)	(.89)	(.36)	(.73)	(.78)
$(U_\tau)(\Delta E)$	−.13321	−.10204	.08013	−.00953	−.04372
	(3.46)	(4.57)	(2.16)	(.20)	(1.16)
$D(\Delta E)$	.01294	−.00049	.03336	−.07712	−.00703
	(.42)	(.02)	(.91)	(1.71)	(.18)
N	2114	2101	1525	1389	1076
R^2	.18	.13	.09	.12	.15
RSS		366.4	294.7	209.4	227.2

[a]Absolute values to t-ratio are shown in parentheses.

form:

$$(4)\quad \ln w_\tau - \ln w_t = B_0 + B_1 E_t + B_2 \Delta S + B_3 \Delta H + B_4 \Delta ST + B_5 U_\tau + B_6 A_t + B_7 M_\tau + B_8 M_t + B_9 (\Delta E)(D) + B_{10}(\Delta S)(\Delta E) + B_{11}(U_\tau)(\Delta E) + B_{12}\Delta E + B_{13} D + e_t.$$

The variables are defined as follows.

E_t = Years of postschool work experience, measured as the sum of weeks divided by 52.

S_t = Years of schooling obtained by year t.

ST_t = A dummy variable set equal to one if the individual is enrolled in school, zero otherwise.

M_t = A dummy variable set equal to one if the individual is married in year t.

D = A dummy variable set equal to one if white, zero if black.

A_t = Age in year t.

U_t = A dummy variable set equal to one if wage is set by collective bargaining.

H_t = Usual weekly hours of work in year t.

The residual e_t is assumed to have all of the classical properties. Equation (4) is differentiated with respect to ΔE in order to obtain a measure of the slope of the average experience-earnings profile. For exam-

TABLE 2—CORRECTED AND OBSERVED RELATIVE WAGE ESTIMATES

	Corrected	Observed
1966–69	1.27	1.26
1967–70	1.87	1.30
1968–71	.59	1.30
1966–71	1.35	1.30

Notes: All corrected wages calculated by the formula

$$w_\tau^c = \overline{w}_\tau + \int_0^{45} (\partial w_\tau / \partial \Delta E) e^{-.1t} \, dt.$$

The observed relative wages are the ratios $\overline{w}_\tau^W / \overline{w}_\tau^B$, while the corrected relative wages are the ratios $w_\tau^{Wc} / w_\tau^{Bc}$ (see equations (2) and (4)).

ple, using (4), the experience-induced wage growth of whites is

$$(5) \quad \partial w_\tau / \partial \Delta E = \overline{w}_\tau (B_{12} + B_{11}\overline{U}_t + B_9 D + B_{10}\Delta\overline{S})|_{D=1}.$$

Estimates of the model for different periods in the *NLS* data set are shown, with Lazear's results, in Table 1.[2] As in Lazear's original regression, the white-experience interaction term is insignificant in the four new regressions.

The corrected relative wage estimates are presented in Table 2. They were calculated under the assumption that $r = .10$ and $N = 45$, and with the wage growth figures and variable means shown in Tables 3 and 4, respectively. What is most important about Table 2 is the instability of the corrected relative wage estimates. This is even more striking when compared to the observed relative wages, which are very close to one another.

Since, as Table 2 confirms, the observed relative wage of white workers is fairly constant over time, it is clear that the wage growth derivative given in equation (5) is responsible for the instability seen in the corrected relative wages. Moreover, estimates of the wage growth model in Table 1 do not

[2] The sample used for the regressions includes only white and black workers and excludes those who reported a wage of zero.

TABLE 3—EXPERIENCE-INDUCED WAGE GROWTH ESTIMATES

	White	Black
1966–69	.126	.096
	(.04)	(.05)
1967–70	.095	−.027
	(.08)	(.08)
1968–71	−.125	.120
	(.10)	(.11)
1966–71	.398	.387
	(.09)	(.10)

Notes: All derivatives calculated according to equation (5) and evaluated at sample means. Standard errors are shown in parentheses.

TABLE 4—MEANS OF VARIABLES USED IN WAGE GROWTH CALCULATIONS

	White	Black
1966–69		
ΔS	1.41	1.27
	(1.30)	(1.30)
U_{69}	.27	.31
	(.44)	(.46)
1967–70		
ΔS	1.11	.95
	(1.21)	(1.20)
U_{70}	.29	.36
	(.45)	(.48)
1968–71		
ΔS	.81	.66
	(1.10)	(1.02)
U_{71}	.28	.34
	(.45)	(.47)
1966–71		
ΔS	1.80	1.53
	(1.80)	(1.71)

Note: Standard deviations are shown in parentheses.

suggest that the model is either unstable or suffers from a multicollinearity problem. If instability or multicollinearity were serious problems here, the coefficient estimates in the four new regressions would differ sharply between regressions, which is apparently not the case. Even between the two regressions whose results are most divergent, the 1966–69 and 1968–71 models, almost all of the coefficient estimates lie within two standard deviations of one another.

The wage growth estimates are sensitive to small changes in either the parameters or the variable means used in calculating them.

TABLE 5—DECOMPOSITION OF DIFFERENCE BETWEEN PERCENTAGE WAGE GROWTH ESTIMATES FROM 1966–69 AND 1968–71 REGRESSIONS

Variable	Means	Coefficients
ΔE	0	$-.027$
$U_T(\Delta E)$	.004	$-.100$
$D(\Delta E)$	0	0
$S(\Delta E)$	$-.055$	.053
Total	$-.051$	$-.074$

Notes: This is a decomposition of $\partial \ln w/\partial \Delta E_{(1)} - \partial \ln w/\partial \Delta E_{(2)}$ where the subscript (1) denotes the percentage growth estimate from the 1966–69 model, while (2) denotes the estimate from the 1968–71 model. The above decomposition is for the black estimate only.

Again considering the most divergent corrected wage ratios, the decomposition in Table 5 shows the factors influencing the different percentage wage growth estimates for black workers. The relatively large share of the difference in percentage wage growth explained by changes in the coefficients is disturbing, particularly since two of the four coefficient estimates used are imprecise. If this example can be considered representative, it demonstrates that Lazear's results are unstable in the sense that small changes in the parameter estimates lead to rather large changes in the wage growth estimates.

II. Concluding Remarks

Although little confidence can be put into the final numbers, Lazear's conclusion that earnings differentials, in a present value sense, have not narrowed seems plausible. More frequent unemployment spells and a shorter average experience horizon for black workers are sufficient explanations—and neither imply an absence of labor market discrimination. But Lazear has relied on a divergence in the slopes of experience-earnings profiles to support his argument, a divergence which is weakly supported by his regression results. Whether, in fact, black-white earnings profiles diverge seems to be an open issue (see Richard Freeman, 1981).

REFERENCES

Freeman, Richard, "Have Black Labor Market Gains Post-1964 Been Permanent or Transitory?," Working Paper No. 751, National Bureau of Economic Research, Cambridge, 1981.

Lazear, Edward, P. "The Narrowing of Black-White Wage Differentials is Illusory," *American Economic Review*, September 1979, *69*, 553–63.

National Longitudinal Survey for Young Men, 1966–71, Ohio State University.

Illusory Wage Differentials: Reply

By EDWARD P. LAZEAR*

With the exception of a few minor points, I do not disagree with Keith Hylton's comment. In fact, I am somewhat puzzled by it because I believe that what Hylton has shown was taken as given in my 1979 paper. That the experience coefficient is not estimated with precision can be inferred directly from the standard error of the wage growth effect. I state explicitly,

> We are quick to add, however, that the difference in estimated *OJT* across the two groups is not large in 1966. In fact, $\bar{H}(1966)|_{D=1} - \bar{H}(1966)|_{D=0} = \$0.55$ (with a standard error of \$0.67) so that the difference is relatively small and the standard error is quite large. The conclusion is then that for this early period, the measured wage differential underestimates the true differential, but the extent of that understatement is questionable. [p. 559]

It seems to me that Hylton has done little more than document that statement. As such, I agree with his evidence.

The important point, however, is that my argument did not rest on the coefficient obtained from one panel, but rather on the difference between the coefficients on the two panels. This point is not addressed in the comment, but should be. Indeed, now that a few years have passed, there is a better and more direct way to test the validity of my original hypothesis. My method was developed to estimate the value of *OJT* or wage growth when the entire life cycle of wages could not be observed. But now, there is considerable evidence on the exact shape of the age-earnings profiles for the individuals who were 14–24-years old in 1966. Much more recent waves of the *National Longitudinal Survey* are currently available. This provides a much clearer test of the hypothesis because divergence of the actual profiles can be examined. There is no need to rely on an early extrapolation as I did when only the first few years of the data were available. To my mind, this provides the basis for a more constructive reexamination of the hypothesis. Hopefully, future work (perhaps by Hylton, myself, or both) will proceed along these lines. In any event, I am pleased that others, too, continue to believe that this is an important issue.

*University of Chicago, Chicago, IL 60637, and National Bureau of Economic Research.

REFERENCES

Hylton, Keith, "Illusory Wage Differentials: Comment," *American Economic Review*, December 1984, *74*, 1124–27.

Lazear, Edward P., "The Narrowing of Black-White Wage Differentials is Illusory," *American Economic Review*, September 1979, *69*, 553–63.

Currency Substitution and Instability in the World Dollar Standard: Comment

By CHRISTOPHER RADCLIFFE, ARTHUR D. WARGA, AND THOMAS D. WILLETT*

Ronald McKinnon has argued that because of international currency substitution, "In general, growth in the world money supply is a better predictor of American price inflation than is U.S. monetary growth" (1982, p. 324). Myron Ross (1983) has shown that regressions for McKinnon's entire sample period (1960–80) do not support this conclusion, but McKinnon and Kong-Yam Tan (1983) have countered that international currency substitution for the United States was really only important from the beginning of the 1970's, and over this period their regressions find the world money supply to explain U.S. prices better than the U.S. money supply.

This empirical correlation may be due in considerable part to the effects of the two major oil price increases during the 1970's (see Willett, 1983) rather than to currency substitution. The appropriate type of tests of the currency substitution hypothesis involves effects on the demand for money, that is, velocity, not directly on prices. The latter would be an equivalent proxy only in a simple classical full-employment world where there was a one-to-one relationship between changes in aggregate demand and the price level.

Currency substitution is a possible explanation of the instability of estimated demand for money functions for the United States during 1970's. Shifts out of the dollar into other currencies would reduce the demand for dollars and increase the velocity of the U.S. monetary aggregates. If sterilization policies keep the U.S. money supply from falling (which is the normal practice), then the increased velocity will lead to an increase in nominal income, and the manner in which this increase is divided between increases in real output and prices may be quite variable. Likewise, a shift toward dollars should reduce the U.S. velocity and nominal *GNP*. Thus, for McKinnon's hypothesis of the dominant role of international currency substitution to be correct, we should find that the velocity of the world money supply with respect to U.S. nominal spending should be more stable than for the velocity of the U.S. monetary aggregates. One way of testing this to see whether the U.S. money supply or the world money supply provides a better explanation of movements in U.S. nominal *GNP* (not just prices). A weaker hypothesis is that foreign money supplies are a significant but only partial substitute for U.S. monetary holdings. To allow for this possibility that foreign money supplies may help explain U.S. *GNP*, but with a different coefficient from the domestic money supply, we tried both in the same equation. Neither the strong nor weak version finds strong support in the data.

I. Empirical Results

The estimated equations test the linkage between growth in money (either M_1US, U.S. domestic money, or M_1W, world money) and growth in *GNP* for the United States. The series on world money was compiled using McKinnon's procedures, a weighted average of domestic money growth rates for ten leading industrial countries, with the weights corresponding to *GNP* in 1970. All tests are on annual data taken from the *International Financial Statistics Yearbook*, 1982. All results presented were estimated using least squares estimation techniques.[1]

*Claremont Center for Economic Policy Studies, Claremont Graduate School and Claremont McKenna College, Claremont, CA, 91711.

[1]Adjustments for first-order serial correlation do not change any of the paper's conclusions. A Durbin procedure was employed to permit cross-equation comparisons of regression diagnostics.

TABLE 1—THE IMPACT OF DOMESTIC VS. "WORLD" MONEY SUPPLY CHANGES ON U.S. *GNP*

Equation	C	M_1US-1	M_1US-2	M_1W-1	M_1W-2	ROW-1	ROW-2	R	SER	D-W	Period
(1)	.038	1.04	−.117					.733	.015	1.34	1958–81
	(5.41)	(7.20)	(−.843)								
(2)	.039	.98	−.276					.796	.012	1.05	1958–69
	(6.13)	(6.58)	(−1.60)								
(3)	.069	.879	−.367					.51	.016	1.61	1970–81
	(3.53)	(3.56)	(−1.53)								
(4)	.011			.778	.156			.498	.021	1.21	1958–81
	(.659)			(4.40)	(.908)						
(5)	.044			.889	.02			.574	.017	2.27	1958–69
	(.191)			(4.05)	(.904)						
(6)	.063			.501	−.075			.177	.02	.786	1970–81
	(2.17)			(2.09)	(2.79)						
(7)	.043	1.02	−.015			.125	−.284	.73	.015	1.12	1958–81
	(3.06)	(5.89)	(−.091)			(.558)	(−1.27)				
(8)	.026	.919	−.266			.263	.045	.768	.012	1.15	1958–69
	(1.36)	(5.38)	(−1.23)			(.966)	(.156)				
(9)	.088	1.12	−.481			−.255	−.215	.456	.016	1.94	1970–81
	(2.78)	(3.20)	(−1.17)			(−.588)	(−.542)				

Table 1 displays results of equations which follow the procedure of McKinnon and Tan, except for the change in the dependent variable from wholesale prices to *GNP*. The original equations estimated by McKinnon using wholesale prices were also estimated in the course of this study, yielding results consistent with McKinnon and Tan. Equations (1)–(3) examine the impact of growth in U.S. domestic money (M_1US, lagged one and two periods) on *GNP*. Equation (1) is estimated over the entire period 1958–81, while equations (2) and (3) are estimated for the subperiods 1958–69, 1970–81, respectively. Equations (4)–(6) repeat the analysis with world money growth (M_1W). For each interval, the domestic money equations perform better than their world money counterparts with respect to R and standard error of the regression (SER).

Particular note should be taken of equation (6) which is far weaker than equation (3). This is the subinterval (1970's) where the McKinnon hypothesis is expected to be most strongly borne out. Equations (7)–(9) present the results of an alternative procedure used by McKinnon and Tan to check the statistical significance of the world money concept. The U.S. money component of world money is subtracted from M_1W to form an incremental variable ROW. This incremental variable is then included in the domestic money regressions (lagged one and two periods) to test the significance of the money supplies of the "rest of the world" in explaining the variation of U.S. *GNP*. The coefficient on this variable is thus allowed to differ from that on M_1US. In the world money equations (4)–(6), both ROW and M_1US implicitly have the same sign, which would restrict the substitution between U.S. and non-U.S. money to be "full," or one to one. This alternative specification is thus a broader test of the currency substitution hypothesis, by allowing the non-U.S. money variable to indicate significant but only partial substitution. Equations (7)–(9) present results over the same intervals as the previous tests. In no equation does the variable ROW add significantly to the explanatory power of M_1US, and in the second subinterval both coefficients on ROW are negative as well as insignificant. This is corroborated by an insignificant F-test on these increments.

Thus we fail to find support for the hypothesis that foreign money supplies have significantly influenced nominal income in the United States. We consider this strong evidence against the view that foreign currency holdings are highly substitutable with dollars. There is still scope for substantial international influences on the U.S. economy

through such traditionally recognized channels as capital and trade flows and direct price effects.[2] We view this as an important area for further empirical research.[3]

[2] Mark Miles (1981) has used another approach to find evidence of statistically significant currency substitution for the United States, although Leroy Laney, Radcliffe, and Willett (forthcoming) have shown that the magnitude of interdependence implied by Miles' results is quite small.

[3] For further discussion, see Willett.

REFERENCES

Laney, Leroy O., Radcliffe, Chris D. and Willett, Thomas D., "International Currency Substitution by Americans Is Not High: A comment on Miles," *Southern Economic Journal*, forthcoming.

McKinnon, Ronald I., "Currency Substitution and Instability in the World Dollar Standard," *American Economic Review*, June 1982, *72*, 320–33.

______ **and Tan, Kong-Yam,** "Currency Substitution and Instability in the World Dollar Standard: Reply," *American Economic Review*, June 1983, *73*, 474–76.

Miles, Marc A., "Currency Substitution: Some Further Results and Conclusions," *Southern Economic Journal*, July 1981, *48*, 78–86.

Ross, Myron H., "Currency Substitution and Instability in the World Dollar Standard: Comment," *American Economic Review*, June 1983, *73*, 473.

Willett, Thomas D., "U.S. Monetary Policy and World Liquidity," *American Economic Review Proceedings*, May 1983, *73*, 43–47.

International Influences on the U.S. Economy: Summary of an Exchange

By RONALD I. MCKINNON, CHRISTOPHER RADCLIFFE, KONG-YAM TAN, ARTHUR D. WARGA, AND THOMAS D. WILLETT*

In "Currency Substitution and Instability in the World Dollar Standard" (1982), Ronald McKinnon hypothesized that shifts in international portfolio preferences for dollar assets—including bonds—destabilized the effective demand for money in the United States. From the dollar depreciation of 1971–73 to the great dollar appreciation of 1981–83, these demand shifts were telegraphed by large changes in the dollar exchange rate against other hard currencies. They signaled sudden inflation in the United States when the dollar was unexpectedly weak, and deflation when the dollar became strong.

In addition, McKinnon argued this ebb and flow in the demand for dollar assets provoked foreign central banks, but not the U.S. Federal Reserve System, to adjust their money growth rates to mitigate these exchange fluctuations. Consequently, since 1970, annual percentage growth in "World" $M1$—the sum of percentage growths in dollars, marks, yen, sterling and so on—fluctuated more than annual money growth in the United States. The resulting international business cycle had a first-order impact on American income and prices.

Other than discussing some suggestive money and price data for ten industrial countries, McKinnon provided no formal econometric testing of his theory. This note summarizes an exchange between McKinnon and Tan (M-T) and Radcliffe, Warga, and Willett (R-W-W) prompted by the latter's 1984 econometric test of McKinnon's hypothesis. Copies of the full exchange, including new econometric work, are available from the authors.[1]

Consider the single-equation econometric technique of explaining U.S. prices or incomes by current or lagged changes in U.S. $M1$. All authors agree that there was a significant deterioration in the fit of this basic monetary equation from the fixed exchange rate period of 1958–69 to the era of floating rates from 1972 to 1982. By itself, American money growth now gives a less satisfactory explanation of cyclical fluctuations in nominal income or prices.

But how can changing international asset preferences be represented statistically in a mixed exchange rate regime of "dirty" floating? One proxy variable is money growth in the rest of the industrial world: $M1^{ROW}$. McKinnon (1982) originally hypothesized that world money inclusive of $M1^{ROW}$ has a stronger impact on American prices than U.S. money by itself. But this conjecture turns out to be true only for American (and world) tradable goods prices—as approximated by the wholesale price index (*WPI*). (See McKinnon-Tan, 1983).

Radcliffe, Warga, and Willett correctly pointed out that the influence of $M1^{ROW}$ is not helpful in predicting changes in U.S. nominal *GNP*. Myles Wallace (1983) showed that domestic price indices, such as the American *CPI* or *GNP* deflator, which have large nontradable components are not well explained by world money. R-W-W's objection is important because monetary variables

*McKinnon and Tan: Department of Economics Stanford, University, Stanford, CA 94305; Radcliffe, Warga, and Willett: Claremont Center for Economic Policy Studies, Claremont Graduate School and Claremont McKenna College, Claremont, CA 91711.

[1]The full exchange may be obtained by writing to Claremont Working Papers, Claremont Center for Economic Policy Studies, The Claremont Colleges, Claremont, CA 91711. Ask for "Currency Substitution and Instability in the World Dollar Standard: Reply and Further Comments," by McKinnon et al. See also Radcliffe, Warga, and Willett (forthcoming).

are often linked to nominal *GNP* through the money-demand function. Whereas the tradeoff between price inflation and real output growth is harder to predict.

McKinnon-Tan disagree with the conclusion that R-W-W's findings are strong evidence against the view that portfolio shifts between dollar and foreign-currency assets effectively destabilize the demand for money in the United States. For shifts in world money to be an adequate proxy for shifts in the demand for dollars, a necessary condition is that exchange rates are predominantly fixed—a simplifying assumption made by McKinnon (1982) in constructing a theoretical model of the international money multiplier.

Under the 1972–82 regime of predominantly floating rates, however, M-T believe that the dollar exchange rate—trade weighted against seventeen other industrial countries —behaves as if it was a good indicator of when the demand for money in the United States shifted. M-T show that the goodness of fit of the basic monetary equation is greatly improved when (changes in) the dollar exchange rate is included as an additional explanatory variable.[2] For a given rate of growth in $M1^{US}$, a 10 percent appreciation of the dollar this year will slow U.S. nominal *GNP* growth by about 2.5 percentage points next year (see also McKinnon, 1984).

Radcliffe, Warga, and Willett agree that the M-T paper represents an important reformulation of the McKinnon hypothesis, and that their original comment therefore does not disconfirm M-T's expanded currency substitution hypothesis. They also argue that McKinnon's work presents a useful counterbalance to the still frequent analysis of developments in U.S. demand for money and inflation which focus almost entirely on domestic considerations (see, for example, Alan Blinder, 1982; Stephen Goldfeld, 1976). They remain skeptical, however, of the importance of international currency substitution for the United States over the past decade. They point out that the new M-T results are also consistent with traditional analysis of the impact effects of exchange rate movements on prices and on output via the Keynesian trade balance mechanism,[3] and that M-T do not present any systematic evidence that international currency substitution was, in fact, a major cause of movements in the dollar.

McKinnon-Tan reply that they believe the effect of the dollar exchange rate on American nominal *GNP* is simply too large to be consistent with a purely "nonmonetary" explanation based on computing the elasticity of response of exports and imports to exchange rate changes. Because the American foreign trade sector is relatively small, they think other influences must be at work. M-T contend that an increase in the dollar exchange rate is also associated with a number of "monetary" factors: a fall in world money growth, a negative wealth effect on dollar debtors throughout the world, and a signal to Americans that future price inflation will be lower.

In order to emphasize nonmonetary influences on fluctuations in U.S. income, R-W-W point out a strong temporal association between several major exchange rate changes (emphasized by M-T) and substantial oil and food price changes. R-W-W present a broad set of new econometric results which suggest the past importance of food

[2]Both sets of authors also point to the difficulties which have been involved with the different and changing definitions of the major U.S. monetary aggregates in recent years. They present new empirical results based on an averaged money supply series for $M1$ reported by the Federal Reserve Bank of St. Louis. They find that this series performs considerably better than the $M1$ series reported by the International Monetary Fund originally used by McKinnon and by R-W-W and find that when changes in the average exchange rate of the dollar are added to their equations they are usually associated with a significant effect on U.S. nominal *GNP* with a one-year lag. The results summarized in this note refer to the St. Louis Fed Measure. The exchange did not tend to be significant in equations with the original $M1$ measures, but great weight should not be placed on this since the equations did so poorly during the floating rate period.

[3]On the price and trade transmission mechanisms, see, for example, Willett and John Mullen (1982) and J. Harold McClure (forthcoming) for empirical analysis. As McClure reminded us, there is also a plausible chain of causation running from exchange rate changes to the demand for money to changes in exchange rates. On the former, see Richard Cooper (1976) and Edward Tower (1975).

and oil prices in the short-run determination of U.S. nominal income, but find that these factors do not rob changes in the exchange rate of significant explanatory power. Foreign money supply development unadjusted for exchange rate changes continue to show little significance for U.S. *GNP* growth.[4] Although in further work, M-T have shown that foreign money growth measured in U.S. dollars has a highly significant impact on American *GNP* (see Tan, 1984; McKinnon, 1984).

All of the participants in this exchange believe that further analysis of the multiple channels of international influences on the U.S. economy should be a priority area for research, both for the analysis of inflation, employment, and output, and for questions of the best definition of American monetary aggregates and the measurement of instability in money demand. R-W-W agree that an expanded test of the McKinnon analysis should include both exchange rate changes and foreign money supply variables. They also agree that exchange rate developments can give important signals for U.S. monetary policy, although R-W-W would be inclined at this point to place less presumptive weight on such changes by themselves. M-T would orient American monetary policy towards smoothing fluctuations in the dollar exchange rate provided that this was consistent with stabilizing aggregate money growth across the hard-currency countries.[5]

[4]See also Henry Goldstein and Stephen Haynes (1984).

[5]A more complete theoretical and empirical analysis of why and how one would implement such a policy is provided in McKinnon (1984) and in Tan (1984).

REFERENCES

Arndt et al., Sven, *Exchange Rates, Trade, and the U.S. Economy*, New York: Ballinger, forthcoming.

Blinder, Alan S., "The Anatomy of Double Digit Inflation in the 1970's," in Robert E. Hall, ed., *Inflation: Causes and Effects*, Chicago: University of Chicago Press, 1982, 261–82.

Cooper, Richard N., "Monetary Theory and Policy in an Open Economy," *Scandinavian Journal of Economics*, No. 2, 1976, *78*, 146–43.

Goldfeld, Stephen, M., "The Case of the Missing Money," *Brookings Papers on Economic Activity*, 3: 1976, 683–730.

Goldstein, Henry N. and Haynes, Stephen E., "A Critical Appraisal of McKinnon's World Money Supply Hypothesis," *American Economic Review*, March 1984, *74*, 217–24.

McClure, J. Harold, Jr., "Dollar Appreciation and the Reagan Disinflation," in S. Arndt et al., eds., *Exchange Rates, Trade, and the U.S. Economy*, forthcoming.

McKinnon, Ronald I., "Currency Substitution and Instability in the World Dollar Standard," *American Economic Review*, June 1982, *72*, 320–33.

______, *An International Standard for Monetary Stabilization*, Washington: Institute for International Economics, 1984.

______ **and Tan, Kong-Yam,** "Currency Substitution and Instability in the World Dollar Standard: Reply," *American Economic Review*, June 1983, *73*, 474–76.

Radcliffe, Christopher, Warga, Arthur D. and Willett, Thomas D., "International Influences on U.S. National Income: Currency Substitution, Exchange Rate Changes, and Commodity Shocks," in S. Arndt et al., eds., *Exchange Rates, Trade, and the U.S. Economy*, forthcoming.

Tan, Kong-Yam, "Flexible Exchange Rates and Interdependence: Empirical Implications for U.S. Monetary Policy," unpublished doctoral dissertation, Stanford University, 1984.

Tower, Edward, "Money Demand and the Terms of Trade," *Weltwirtschaftliches Archiv*, December 1975, *111*, 623–33.

Wallace, Myles, "Currency Substitution and Instability in the World Dollar Standard: Additional Evidence," unpublished, Department of Economics, Clemson University, June 1983.

Willett, Thomas D., "U.S. Monetary Policy and World Liquidity," *American Economic Review Proceedings*, May 1983, *73*, 43–47.

______ **and Mullen, John,** "The Effects of Alternative International Monetary Systems on Inflationary Biases and Macroeconomic Discipline," in Raymond Lombra and Willard Witte, eds., *Political Economy of International and Domestic Monetary relations*, Ames: Iowa State University Press, 1982, 143–46.

Why Do Companies Pay Dividends?: Comment

By GARY A. BORTZ AND JOHN P. RUST*

In a recent paper in this *Review* (1983), Martin Feldstein and Jerry Green present a model of stock market equilibrium which seems to solve the "dividend puzzle." The model shows that firms maximize their market value by paying a determinant, positive level of dividends despite the substantial tax penalty on dividends relative to capital gains. Such a result is potentially important because it is consistent with Fischer Black and Myron Scholes' (1974) "supply effect" argument that changes in dividend policy are irrelevant, yet at the same time it explains why firms maintain a determinant dividend payout—something the supply effect argument fails to do.[1] Feldstein and Green attribute their result to "the combination of the conflicting preferences of shareholders in different tax brackets and their desire for portfolio diversification in the face of uncertainty that together cause all firms to pay dividends" (p. 29).

Unfortunately, this appealing explanation is incorrect. Feldstein-Green's result is an artifact of the assumption that firms face a tradeoff between dividends and investment. Implicit in their model is the constraint that firms cannot borrow or issue new shares to finance dividend payments. If we remove this constraint and allow firms to raise new capital for dividend payments, the Feldstein-Green result evaporates. In fact, the completely opposite result obtains; namely, firm value is maximized when dividends are zero and changes in dividend policy necessarily affect share price in equilibrium. The explanation of this result is quite straightforward. *When dividend and investment policy are linked, as they are in the Feldstein-Green model, changes in dividend policy imply changes in investment policy which affects the risk-return profile of the firm.* Therefore it should come as no surprise that a determinant level of dividends will maximize market value in such a model, even in the presence of taxation of dividend income. Dividend policy becomes a proxy for investment policy, enabling consumers to obtain a better balance between risk and return in their portfolios. When investment policy is fixed, consumers must allocate their portfolios between riskless bonds and risky stocks whose random return per dollar is exogenously fixed given the current level of investment. When dividends are introduced, the level of investment is no longer independently determined. By varying dividend policy, the firm changes the level of investment altering the return distribution of the stock, providing an advantageous extra degree of freedom for balancing risk and return in consumer portfolios. With risk-averse investors it will generally be optimal to use this extra degree of freedom, which explains the existence of positive dividends. Dividend policy is "relevant" in such a model only because investment policy is relevant, but this is a point that few of us would find surprising.

The appropriate way to study dividend policy is to isolate it from investment policy. This can be done quite simply by allowing the firm to finance dividend payments through issue of new stocks or bonds, leaving the level of investment unaffected. Once we

*The First Boston Corporation, Park Avenue Plaza, New York, NY 10055 and University of Wisconsin, 1180 Observatory Drive, Madison, WI 53706, respectively.

[1] The supply effect argument of Black and Scholes states that if a capital market is in supply equilibrium then dividend policy is necessarily irrelevant. They say: "If corporations are generally aware of the demands of other investors for low dividend yields, they will adjust their dividend policies to supply the level of yield that are most in demand at any particular time. After equilibrium is reached no firm will be able to affect its share price by changing its dividend policy" (p. 1). The principal problem with this explanation is that any pattern of payouts could be consistent with equilibrium. This is the gap which needs to be filled. As a caveat, note that Barr Rosenberg and V. Marathe (1980) and Robert Litzenberger and Krishna Ramaswamy (1979) have found evidence inconsistent with the supply effect argument.

do so it follows immediately that the optimal dividend payment is zero due to the tax penalty on dividends. In the absence of taxes changes in dividends have no effect on share prices since investors can simply rebalance their portfolios purchasing more shares of stock to offset the dividend income. It follows that any portfolio obtainable when dividends are zero can be duplicated by appropriate adjustments when dividends are positive. This is the essence of the Merton Miller and Franco Modigliani (1961) dividend irrelevance theorem. However with dividends and taxes, investors cannot maintain their desired portfolios without incurring tax liabilities. It follows that increases in dividends decrease investors' expected utility, implying that in a competitive stock market equilibrium dividend payments of value-maximizing firms are zero.

The tradeoff between dividends and investment which Feldstein-Green implicitly assume appears to have little empirical validity: with access to large, efficient capital markets few publicly traded firms would find it necessary to reduce investment in order to pay regular dividends to shareholders. We conclude that the Feldstein-Green paper has not brought us any closer to a solution of the "dividend puzzle." Instead, it provides another illustration of a basic mistake warned about in finance textbooks: "Every financial decision is a tradeoff simply because sources and uses of funds has to balance. It is critically important to identify the tradeoff before thinking about any financial problem" (Stewart Myers, 1976, p. 9).

REFERENCES

Black, Fischer and Scholes, Myron, "The Effects of Dividend Yields and Dividend Policy on Common Stock Prices and Returns," *Journal of Financial Economics*, May 1974, *1*, 1–22.

Feldstein, Martin and Green, Jerry, "Why Do Companies Pay Dividends?," *American Economic Review*, March 1983, *73*, 17–30.

Litzenberger, Robert and Ramaswamy, Krishna, "The Effect of Personal Taxes and Dividends on Capital Asset Prices: Theory and Empirical Evidence," *Journal of Financial Economics*, 1979, *2*, 163–96.

Miller, Merton and Modigliani, Franco, "Dividend Policy, Growth, and the Valuation of Shares," *Journal of Business*, October 1961, *4*, 411–33.

Myers, Stewart, *Modern Developments in Financial Management*, Hinsdale: Dryden Press, 1976.

Rosenberg, Barr, and Marathe, V., "Tests of Capital Asset Pricing Hypotheses," *Journal of Financial Research*, 1980, *4*, 121–35.

Why Do Companies Pay Dividends?: Comment

By JOEL HASBROUCK AND IRWIN FRIEND*

In a recent article in this *Review*, Martin Feldstein and Jerry Green (F-G) note, "the nearly universal policy of paying substantial dividends is the primary puzzle in the economics of corporate finance" (1983, p. 17), and propose a new theoretical explanation for dividend payments in spite of unfavorable taxation. The payment of dividends, according to Feldstein-Green, reflects appropriate management response to the desire for a diversified investment portfolio by risk-averse investors who regard each firm's return as "both unique and uncertain." Unfortunately, their argument is flawed by several shortcomings which, in our view, substantially reduce its plausibility.

The basic but by no means only important deficiency in the model on which the F-G analysis and conclusions are based lies in its treatment of risk-free assets. In their model, investors can obtain the risk-free asset only by investing dividend income. We suggest that relaxing this assumption will greatly mitigate the impetus for dividend payout. A second questionable assumption in the F-G model is that while corporations provide the only funds available to acquire net risk-free assets, they cannot hold them in their own portfolios. We shall show that relaxation of this restriction will lead to total retention. F-G do state at the end of their paper that one worthwhile extension of their model "would be to recognize...that corporations as well as investors can earn the risk-free return" (p. 29). However, they surmise that the link between dividends and real corporate investment that is implicit in their present model, though weakened, would persist. We show that this surmise is incorrect.

Still another serious deficiency in the F-G model is their misspecification of the quadratic utility function, which they say they are using. While we hold no special brief for the quadratic utility function, we shall demonstrate that not only do F-G misspecify it, but more importantly, the utility function they use does not seem to conform to any reasonable decision rule. Finally, we shall discuss briefly other significant problems in the F-G analysis which they do not recognize, at least explicitly.

I. The Role of the Risk-Free Asset

In the F-G model, an investor desiring to invest in the risk-free asset is crucially affected by the corporate sector's dividend policy for two principal reasons: dividend payments are the only source of funds investable at individual discretion; and, corporate holdings of the risk-free asset are forbidden. Concerning the first point, note that aggregate holdings of the risk-free asset in their model derive exclusively from dividend payments used to purchase a risk-free security that is presumably supplied in a neutral and acquiescent fashion by the government. Even neglecting the reasonableness of the supply-side assumptions, it is clear that, in reality, individuals have many sources of income that may readily be diverted into the risk-free asset. The importance of this force compelling dividend payments therefore seems questionable.

Concerning the second point, we shall demonstrate that the F-G model does not appear to support positive dividend payments once corporations are allowed to hold risk-free assets.[1] We shall for expository purposes greatly simplify their model but still retain all the essential features for investigat-

*The Wharton School, University of Pennsylvania, Philadelphia, PA 19104.

[1]It has been pointed out that another factor leading to dividend payments in the case of non-publicly traded firms is the dim view taken by the IRS of excessive cash holdings. This consideration does not apply, however, to publicly traded firms, who distribute the bulk of dividend payments and on whom the dividend controversy primarily centers.

ing the net balance of the positive and negative effects of diversification and differential taxation respectively in dividend payout. We shall retain the F-G assumptions of two firms, each with one share outstanding and earnings of one dollar, but now assume only one investor, no initial net supply of risk-free assets, and with both firms equal in expected after-tax return (i.e., $r_1^e = r_2^e = r^e$ in the F-G terminology), in total risk ($\sigma_1 = \sigma_2 = \sigma$), and with a zero correlation between the returns of the two firms ($\sigma_{12} = 0$). It should be noted that this last assumption ($\sigma_{12} = 0$) enhances the potential of diversification for risk reduction, thus making the best possible case for the F-G position. Under their assumptions, obviously the optimal dividend payment d will be the same for the two firms.

With these simplifying assumptions, the F-G one-period model would imply

$$(1)\quad E[W] = 2dR(1-\theta) + 2(1-d)[fR + (1-f)r^e],$$

where W is terminal wealth, R is the risk-free rate before personal income taxes, θ is the effective personal tax rate, and f is the proportion (or amount) of earnings which the corporations invest in risk-free assets. For the moment, we shall assume that R is the rate of return received by both individuals and corporate investors on assets which are both tax free and risk free. Under these assumptions, it is obvious that the variance of W becomes

$$(2)\quad \mathrm{Var}(W) = 4(1-d)^2(1-f)^2\sigma^2.$$

Then, holding risk constant at any level, say $(1-d)(1-f) = k$, we can write

$$(3)\quad E[W] = 2dR(1-\theta) + 2(1-d)R - 2kR + 2kr^e$$

To obtain the effect of dividend payments on wealth for given risk, we compute

$$(4)\quad dE[U(W)]/dd = dE[W]/dd = 2[R(1-\theta) - R] = -2\theta R < 0,$$

so that zero dividends will be paid out.

Dropping the simplifying assumption that investors and corporations both invest in tax-free, risk-free assets and assuming instead that they invest in taxable risk-free assets, with θ_p the personal tax rate and θ_c the somewhat higher corporate tax rate, it is easy to show that

$$(5)\quad dE(W)/dd = 2\{(1-\theta_p)[1+(R-1)(1-\theta_p)] - [1+(R-1)(1-\theta_c)]\},$$

where R is now a before-tax rate of return, and as before, risk is held constant by maintaining $(1-d)(1-f) = k$. This expression is clearly negative for any reasonable value of the parameters, including those used by F-G.

II. The F-G Utility Function

Individual behavior in the F-G economy is governed by expected utility maximization. They assume a utility function quadratic in wealth and claim that it leads to an expected utility function of the form $E[U(W)] = E[W] - (\gamma/2)\mathrm{Var}(W)$, a form linear in mean and variance (henceforth, "linear expected utility"). From this specification follow share-demand functions linear in share price and invariant to risk-free asset holdings. As these utility assumptions constitute a key component of their model and contribute in great measure to its analytical tractability, it is necessary to examine their economic validity and reasonableness.

First, a quadratic utility function (say $U(W) = W - (b/2)W^2$ restricted to the domain $W < 1/b$) will lead to an expected utility of the form $E[U(W)] = E[W] - (b/2)(\mathrm{Var}(W) + E[W]^2)$, a form linear in variance, but quadratic in the mean. From this observation, we may move to a broader question: even though a quadratic utility function does not lead to a linear expected utility, is there some (possibly indeterminate) utility function that does? For suitable restrictions on the types of outcomes considered, it may be possible to construct such utility functions. The general nature of these restrictions is far from obvious, however. The special case of normally distributed out-

comes is of particular importance. In this case, John Chipman (1973) has shown that when the expected utility $E[U(w)]$ is written as a function of mean and variance (μ and σ^2), existence of an underlying utility function is contingent on (among other things) a boundary condition of the form $(1/\sigma)\,dE[U]/d\sigma^2 = d^2E[U]/d\mu^2$. Expected utility linear in mean and variance violates this condition.

We now consider a still more general question: does an expected utility function linear in mean and variance constitute a reasonable decision rule for selecting among uncertain prospects? The answer appears to be in the negative, since such an expected utility is inconsistent with the principles of stochastic dominance.

The concept of stochastic dominance was developed to clarify those properties of random gambles which lead to unambiguous ranking under minimal assumptions about preferences.[2] One risky outcome is said to dominate another, in the first-order stochastic dominance (*FSD*) sense, if it offers enhanced probabilities of higher payoffs. Formally, the distribution function of the dominating gamble's outcomes lies on or below that of the dominated gamble, which implies that the probability mass for the dominating gamble lies to the right of the dominated. Josef Hadar and William Russell (1969) show the equivalence of the following two propositions: (*i*) gamble A dominates (*FSD*) gamble B; and (*ii*) all expected utility maximizers with positive marginal utility of wealth will prefer A to B. As shown in the Appendix, gambles may be constructed in which the ranking obtained under stochastic dominance is reversed using an expected utility linear in mean and variance. For a general utility function, such behavior is tantamount to marginal utility that is not everywhere positive. When employing quadratic utility, the aforementioned wealth restriction is used to exclude the region of negative marginal utility. It is unclear what similar restrictions might suffice to exclude such situations in the linear expected utility case.

In the F-G model, a key determinant of the optimal dividend payout level is the risk-aversion parameter γ, and it would accordingly be highly desirable to obtain reasonable estimates for this parameter. A customary measure of risk aversion is the coefficient of relative risk aversion, defined for a utility function $U(W)$ as $C = -WU''/U'$. The empirical work has for the most part focused on this parameter, and while not completely definitive, suggests that C is constant across wealth levels and in the range 2–6.[3] The preceding discussion has established, however, the general unsuitability of linear expected utility and the tenuousness of its connections to more reasonable constructs. In particular, it does not seem possible to directly relate the linear expected utility risk aversion parameter γ to the usual measure C, and it therefore seems that there is no convenient way to obtain a value for the former from empirical estimates of the latter.[4]

Given the difficulties inherent in the linear expected utility cited above, it is fair to ask how the F-G model might behave under more reasonable utility assumptions. The answer to this question is unclear. The assumption of linear expected utility greatly expedites the analysis by leading to linear

[2]The initial discussion of stochastic dominance is in Hadar and Russell.

[3]See Friend and Marshall Blume (1975), our paper (1982), and Sanford Grossman and Robert Shiller (1981).

[4]A coefficient of relative risk aversion for the linear expected utility function can only be defined in an *ad hoc* fashion by hypothesizing a utility function conditional on knowledge of the expected outcome. Since $\mathrm{Var}(W) = E[W^2] - E[W]^2$, the F-G expected utility function may be written as $E[U(W)] = E[W] - (\gamma/2)[E[W^2] - E[W]^2]$. The function which might logically be viewed as underlying this is $U(W) = W - (\gamma/2)[W^2 - E[W]^2]$, although since this depends on prior knowledge of the expected value of the outcome, it is not a legitimate utility function. Its coefficient of relative risk aversion may be computed as $C = \gamma W/(1-\gamma W)$. In the F-G model, γ appears only as the product $\gamma\sigma^2$, for which a value of 1.87 is assumed in order to lead to a dividend payout of 0.8 and a stock price of 0.78. If each investor's initial endowment is assumed to consist of one-half of each of the two firm's shares, the initial wealth is also 0.78. A realistic estimate for σ^2 follows from noting that the annual return variance for New York Stock Exchange stocks as a whole is approximately 0.04, implying a γ of about 47. Using these values for initial wealth and γ leads to a negative coefficient of relative risk aversion.

share-demand functions. The demand functions derived from a true quadratic utility are not linear in share price and in consequence the resulting system is not easily solved. We acknowledge, however, that the risk-return tradeoff inherent in plausible alternative utility functions may quite possibly lead to an interior solution in which dividends are paid out, maintaining all other assumptions of the model, but this has yet to be demonstrated.

Additional problems with the F-G model, although of lesser importance than those addressed above, stem from the competitive structure of the economy. One troublesome feature of the F-G numerical example is that the current share price seems to be greater than the expected value of the future wealth accruing to the share, at least from the viewpoint of the household investor. Using the F-G parameter values, the share price is 0.78, but the expected value of terminal wealth for the household investor is $dR(1-\theta)+(1-d)r^e=.8(1.1)(1-.5)+.2(1.3)=.70$. This state of affairs could not constitute an optimum for a price-taking investor. The source of this valuation discrepancy is unclear, but appears to arise from the fact that the two firms are not taking investors' implicit required rate of return as exogenous. Robert Merton and Marti Subrahmanyam (1974) show that such a violation of the perfect competition assumption will lead to non-Pareto optimal allocations. They also note that monopoly access to projects (an F-G assumption) will generally lead to investment levels in excess of those socially optimal.

The consequences of large numbers of firms have already been noted by F-G, but the behavior of their model with large numbers of investors is also of some interest. The essential features of the problem may be demonstrated within the framework of the F-G model by assuming an investor clientele consisting of a number of identical households. In such a model, the optimal retention ratio may be shown to depend positively on the number of investors.[5]

III. Conclusion

The conclusion we draw from our analysis of the Feldstein-Green model is that the desire for diversification by risk-averse investors does not explain to any substantial extent why corporations pay dividends in spite of unfavorable taxation.

Appendix

Here it is shown that the linear expected utility rule is inconsistent with the principles of first-order stochastic dominance. Consider gambles in which the wealth outcome is distributed uniformly on the interval (a, b). For fixed a, an increase in b will lead to a new gamble which dominates (FSD) the original (the new distribution function is shifted to the right). We now show that it is possible to construct for the linear utility function uniform gambles for which an increase in the upper endpoint will lead to diminished expected utility.

For a random wealth uniformly distributed on (a, b), the mean and variance are $E(W)=(a+b)/2$ and $\mathrm{Var}(W)=(b-a)^2/12$ (see Harold Larson, 1974). Substituting these expressions into the linear expected utility function and differentiating with respect to the upper endpoint yields

$$dE[U(W)]/db=\tfrac{1}{2}-\gamma(b-a)/12.$$

It is clearly possible to select a and b so that this expression is negative.

[5]If n is taken as the number of identical households, the F-G aggregate feasibility constraint (12) becomes $ns_h=1$. Assuming for simplicity that $\sigma_1^2=\sigma_2^2=\sigma^2$ and $\sigma_{12}=0$, F-G (14) becomes for each firm

$$p_i=1/R[R(1-\theta)d_i+(1-d_i)r_i^e] -(\gamma/nR)(1-d_i)^2\sigma^2,$$

leading to

$$(1-d_i^*)=(n/\gamma)\cdot(1/2\sigma^2)\cdot[r_i^e-R(1-\theta)].$$

REFERENCES

Chipman, John S., "The Ordering of Portfolios in Terms of Mean and Variance," *Review of Economic Studies*, April 1973, *40*, 167–90.

Feldstein, Martin and Green, Jerry, "Why Do Companies Pay Dividends?," *American Economic Review*, March 1983, *73*, 17–30.

Friend, Irwin and Blume, Marshall, "The Demand for Risky Assets," *American Economic Review*, December 1975, *65*, 900–23.

______ **and Hasbrouck, Joel**, "The Effect of Inflation on the Profitability and Valuation of U.S. Corporations," in Marshall Sarnat and Giorgio P. Szego, eds., *Savings, Investment and Capital Markets in an Inflationary Economy*, New York, 1982.

Grossman, Sanford J. and Shiller, Robert J., "The Determinants of the Variability of Stock Market Prices," *American Economic Review*, March 1981, *71*, 222–27.

Hadar, Josef and Russell, William R., "Rules for Ordering Uncertain Prospects," *American Economic Review*, March 1969, *59*, 25–34.

Larson, Harold J., *An Introduction to Probability Theory and Statistical Inference*, New York, 1974.

Merton, Robert C. and Subrahmanyam, Marti G., "The Optimality of a Competitive Stock Market," *Bell Journal of Economics*, Spring 1974, *5*, 145–70.

Why Do Companies Pay Dividends?: Comment

By ODED H. SARIG*

In a recent article in this *Review* (1983), Martin Feldstein and Jerry Green (F-G) posit a model in which "the combination of the conflicting preferences of shareholders in different tax brackets and their desire for portfolio diversification in the face of uncertainty...together cause all firms to pay dividends..." (p. 29). In the model, the double taxation cost of riskless dividends is offset by the gain from the reduction in the investors' risk exposure. F-G derive (under some mild restrictions) an optimal dividend policy that corresponds to a choice of a portfolio of (costly) safe dividends and risky firms' returns.

However, F-G fail to recognize the possibility that firms *internally* reduce the risk of their assets by investing some of their (initial) value in the riskless asset (as opposed to investing its entirety in the risky technology.) If such internal riskless opportunities are availed, the value of dividends as a means for risk reduction disappears.

In terms of F-G's model, this means a substitution of the ith firm dividend (d_i) for an equal investment in the riskless asset. This makes the end-of-period wealth of the households in the model (the equation numbers used match F-G numbering):

$$(2') \quad W_H = R[S_{H1}d_1 + S_{H2}d_2] + Rz_H + S_{H1}(1-d_1)r_1 + S_{H2}(1-d_2)r_2.$$

Note that it is assumed that the risk-free rate of return is the same for the households and the firms. This lack of friction in the financial markets is in line with F-G assumptions.

Adjusting accordingly $E(W_H)$ and solving for **P** as in F-G we get

$$(14') \quad P' = \frac{1}{2R}\begin{bmatrix} R\cdot 2\cdot d_1 + 2(1-d_1)r_1^e \\ R\cdot 2\cdot d_2 + 2(1-d_2)r_2^e \end{bmatrix} - \frac{\gamma}{2R}\begin{bmatrix} (1-d_1)^2\sigma_{11} + (1-d_1)(1-d_2)\sigma_{12} \\ (1-d_2)^2\sigma_{22} + (1-d_1)(1-d_2)\sigma_{12} \end{bmatrix}.$$

Differently stated,

$$(14'') \quad P_i' = d_i + \frac{1}{R}\Big[(1-d_i)r_i^e - \frac{\gamma}{2R}\big[(1-d_i)^2\sigma_{ii} + (1-d_1)(1-d_2)\sigma_{12}\big]\Big].$$

That is the value of each firm is the sum of the value of the riskless investment and the (market) value of the risky investment.

Comparing equation (14′) with F-G equation (14) yields $P_i' - P_i = (\theta/2)d_i > 0$ for $\theta, d_i > 0$. Therefore, a management that seeks to maximize its shareholders' wealth should choose to reduce their risk exposure internally (rather than by means of dividends.)

Note that d_i retains its nature as a decision variable with the optimal investment in the riskless asset satisfying

$$(16') \quad (1-d_i^*) = \frac{(r_i^e - R)}{\gamma\sigma_{ii}} - \frac{(1-d_j)\sigma_{12}}{2\sigma_{ii}}.$$

Evidently, with internal investment in the riskless asset, the effect of corporate taxes should be explicitly considered. However, this is a different line of argument that is not within the scope of F-G's model.

REFERENCE

Feldstein, Martin and Green, Jerry, "Why Do Companies Pay Dividends?," *American Economic Review*, March 1983, *73*, 17–30.

*Columbia Business School, 424 Uris Hall, New York, NY 10027.

NOTES

The ninety-seventh annual meeting of the American Economic Association will be held in Dallas, Texas, December 28–30, 1984.

The Professional Placement Service will be located at the Convention Center. It will be open from 10:00 A.M. to 5:00 P.M., December 27; 9:00 A.M. to 5 P.M., December 28–29, and 9:00 A.M. to 12:00 noon, December 30.

Call for Papers for the 1985 AEA Meetings

Members wishing to give papers or make suggestions for the program for the meetings to be held in New York, NY, December 28–30, 1985, are invited to send their ideas to Alice M. Rivlin, The Brookings Institution, 1775 Mass. Ave., NW, Washington, D.C. 20036. As in the past, the sessions sponsored by the American Economic Association will be of two types: invited papers and contributed papers. Most of the invited papers will be published in the *Papers and Proceedings* issue to appear May 1986, the contributed papers will not be so published. Suggestions of topics or proposals for papers for the invited sessions should be submitted as soon as possible. Abstracts of proposed contributed papers must be received no later than February 1, 1985. Econometric studies or highly mathematical papers are not appropriate for the sessions sponsored by the American Economic Association. However, members wishing to present such papers may submit their proposals or abstracts to the Econometric Society which meets with the American Economic Association, and normally schedules a number of sessions involving such papers.

The American Economic Association discontinued its discount agreement with the Hertz Corporation on August 1, 1984.

Extension of deadline: The National Council for Soviet and East European Research, 1755 Mass. Ave., NW, Washington, D.C., extends the deadline for submission of research proposals from November 1, 1984, to March 31, 1985 (last postmarked date). See the earlier notice in this *Review*, June 1984, p. 543.

Correction: The third World Congress for Soviet and East European Studies will be held at the Sheraton Washington Hotel, Washington, D.C., October 30–November 4, *1985*: hosted by the American Association of Slavic Studies; and cosponsored by the Association and the International Committee for Soviet and East European Studies. The Program Committee Chair is Professor Donald W. Treadgold, School of International Studies, Seattle, WA 98195.

The 1985 meetings of the Southern Regional Science Association will be held May 9–11, at the Washington Hotel, Washington, D.C.

The International Association of Agricultural Economists will hold its nineteenth conference, Agriculture in a Turbulent World Economy, August 26–September 4, 1985, in Malaga, Spain.

The fourth Annual Presidential Conference, John F. Kennedy: The New Frontier, will be held March 28–30, 1985, at the Hofstra University Cultural Center, Hofstra University, Hempstead, NY.

The eleventh Eastern Economic Association meeting will be held in Pittsburgh, March 21–23, 1985.

The International Association of Energy Economists will hold the seventh Annual International Conference, Energy and Economy—Global Interdependencies, in Bonn, FRG, June 3–5, 1985.

A conference, Biased Selection in Health Care Markets, will be held at the University of California-Berkeley, April 11–12, 1985. The conference is supported by UCB, SRI International, Project HOPE, and Medical College of Virginia, VCU.

The third CESOM IRM International Conference on Distribution will be held in Angera, Italy, April 19–20, 1985. It is cosponsored by New York University and L. Bocconi University.

The American Statistical Association/Census Bureau Research Program seeks applications for Fellowships and Associateships beginning September 1, 1985. Positions will be at the Census Bureau for one year or less. Requirements for Fellows: Ph.D. and research record in relevant field (economics, statistics, demography, sociology, business, computer science); for Associates: at least two years graduate study in relevant field. Salaries will be commensurate with qualifications and experience; fringe benefits and travel allowance will be provided. Apply by January 1, 1985 for Fellows; February

15, 1985 for Associates. Contact Dr. J. R. Norsworthy, Chief, Center for Economic Studies, Rm 3442-3, Bureau of the Census, Washington, D.C. (telephone 301+763-2337).

The nineteenth International Atlantic Economic Conference will be held in Rome, Italy, March 9-16, 1985. The theme is Searching for New Ideas. To present a paper, send two copies of a 500-word summary with a cover sheet giving name, affiliation, telephone number of author(s), and appropriate *JEL* category, and $35 (U.S.) submission fee. Prospective discussants and chairs should send same information (not fee). Contact John M. Virgo, International Program Chairman, AEC, Southern Illinois University, Box 101, Edwardsville, IL 62026-1001.

The International Health Economics and Management Conference will be held in conjunction with the above conference. The same dates and guidelines apply for paper presentation ($50 fee). Accepted authors to provide personal expenses and registration fee. Contact John M. Virgo (above address, noting International Health...).

The first annual meeting of the Economists of New Jersey will be held March 30, 1985, at Trenton State College. Sessions will deal with a broad spectrum of economic topics in addition to issues relating to the State of New Jersey. For information and participation forms, contact Annette Meyer of the ENJ Steering Committee, Department of Economics, Trenton State College, Trenton, NJ 08625.

Call for Papers: The University of Cincinnati and the School of Economics at the Autonomous University of Nuevo Leon, Monterrey, Mexico are planning an International Conference on the Economics of Environmental Protection in Mexico in late spring 1985. Papers on economic aspects of environmental protection and resource economics in developing countries in general and in Mexico in particular are invited. Submit a one-page abstract by January 31, 1985. Authors from Mexico and Latin American should contact Lic. Manuel Silos M., Director, Facultad de Economia, Universidad Autonoma de Nueva Leon, Apartado Postal 288, Monterrey, N.L. 64000, Mexico; all others contact Professor Haynes C. Goddard, Department of Economics (371), University of Cincinnati, Cincinnati, OH 45221.

Call for Papers: The fifth World Congress of the Econometric Society will be held at Massachusetts Institute of Technology, August 17-24, 1985. To present papers contact Truman Bewley, Chairman of the Program Committee, Cowles Foundation, Box 2125 Yale Station, New Haven, CT 06520. Applications must include names, addresses, and abstracts. The deadline is January 31, 1985. The Congress is open to all economists. Limited travel funds will be available for those presenting papers.

Call for Papers: The Southern Economic Association will hold its annual meeting at the Hyatt Regency Dallas, Dallas, TX, November 24-26, 1985. Submit two copies of one-page abstracts and cover sheet and $10 fee by February 1, 1985. For details, see the Call for Papers in the October 1984 *Southern Economic Journal*, or write to Dr. Joseph M. Jadlow, College of Business Administration, Oklahoma State University, Stillwater, OK 74078.

Call for Papers: The Military Applications Section of the Operations Research Society of America is offering a $500 Koopmans Prize for the best military operations research paper published in calendar year 1984. "Published" is defined very loosely to mean anything from appearance in or acceptance by a referred journal to appearance as an approved in-house working paper. Entrants do not have to be a member of the Society. Send four copies by January 31, 1985, to Charles Dale, U.S. Army Research Institute, Att: PERI-RG, 5001 Eisenhower Avenue, Alexandria, VA 22333.

The Emory University Law and Economics Center will sponsor the LEC Law Institute for Economists, June 23-July 6, 1985 (date subject to change), in Hanover, NH. It is a program in law, not economics. All participants receive private rooms, most meals, books, and tuition. Application requires a current curriculum vita and a statement that you plan to attend the entire program if accepted. Contact Marc Hoberman, Program Administrator, Emory University, Law and Economics Center, Atlanta, GA 30322 (telephone 404+329-5771). Deadline is January 25, 1985.

The Division of Science Resources Studies of the National Science Foundation announces the continuation of its Program for the Analysis of Science Resources. Proposals may be submitted at any time; however, they should be received by January 18, 1985, for funding in the 1985 fiscal year. Most awards are expected to be for one to two person-years of effort. Areas of interest and types of analyses sought are enumerated in the Announcement of the Program which may be obtained from the Division of Science Resources Studies, NSF, 1800 G Street, NW, Washington, D.C. 20550.

An Advanced Studies Institute in Urban and Regional Science for Developing Countries will be held in Calcutta, December 6-20, 1985, at the Indian Statistical

Institute. Lectures, workshops, and seminars will discuss problems concerning urban development and demography, energy and the environment, decentralization, health and social services, housing, and transportation. There is possible travel support for participants from developing countries. For further details, contact Professor Manas Chatterji, School of Management, State University of New York, Binghamton, NY 13901.

The Arne Ryde Symposium will be held August 26–27, 1985, at the University of Lund. The theme is Incentive Mechanisms and Incentive Problems in Some Major Economic Systems. To present papers, submit abstracts by April 1, 1985. Participants will receive free room and board at the Frostavallen Conference Hotel, and partial travel expenses. Contact Claudio Vedovato, Department of Economics, Box 5137, S-220 05 Lund, Sweden.

Harvard Law School offers fellowships to college and university of teachers in the social sciences and humanities to enable them to study fundamental techniques, concepts, and aims of law, so that, in their teaching and research, they will be better able to use legal materials and legal insights which are relevant to their own disciplines. Further information may be obtained from the Chairman, Committee on Liberal Arts Fellowships in Law, Harvard Law School, Cambridge, MA 02138.

The Board of Governors of the Federal Reserve Minority Doctoral Fellowship Program provides financial support for graduate education leading to the Ph.D. to members of racial or ethnic minorities who have demonstrated high potential in economics or finance. Fellowships will be awarded only to U.S. citizens of racial and ethnic minority groups that have traditionally been underrepresented in these fields. Recipients are selected competitively on the basis of scholastic achievement, faculty recommendations, financial need, and evidence of potential for successful completion of doctoral work in areas of economics or finance of concern to the Federal Reserve Board. Some preference will be given to those who have already completed one or more years of their graduate education, but first-year students are also eligible. Fellowship awards typically include $450 for books, fees, and supplies; a $550 stipend for each month of the academic year and/or tuition; and the option of summer employment at the Board. The deadline for applications for the 1985–86 school year is February 1, 1985. For further information, write to the Board of Governors of the Federal Reserve System, Minority Doctoral Fellowship Program, Board Equal Employment Opportunity Program Office, 20th & C Streets, NW, Washington, D.C. 20551.

The Center for Dewey Studies and the John Dewey Foundation sponsor academic-year fellowships for research. Awards are available to senior scholars in various disciplines interested in exploring issues and problems related to John Dewey's philosophical concerns. The grants normally extend for one year and range from $10,000 to $20,000. They may cover the costs of salary, benefits, travel, secretarial or research support, or research materials. An applicant may hold a concurrent sabbatical salary or small supplementary grant, but not another major fellowship. Application forms are available from The Center for Dewey Studies, Southern Illinois University, Carbondale, IL 62901 (telephone 618+453-2629). Application deadline for the 1985–86 academic year is February 15, 1985. Awards will be announced April 1, 1985.

Pre- or postdoctoral graduate students are invited to submit papers for the 1985 Dorothy S. Thomas Award competition, established by the Population Association of America, in fields of internal migration or the interrelationships among social, economic, and demographic variables. The award is $1,500. For further details, contact Dr. Joan M. Herold, Division of Reproductive Health, Bldg 1, Rm 4409, Centers for Disease Control, Atlanta, GA 30333. The deadline is January 9, 1985.

The P.W.S. Andrews Memorial Prize is awarded annually for an essay by a young scholar (under the age of 30 or within eight years of taking his/her degree) in the general field of Industrial Economics and the Theory of the Firm, broadly interpreted. The prize is £300 and the essay will normally be published in the *Journal of Industrial Economics*. Submissions should be a work of original research by the candidate only, not previously published nor awarded any other prize. The deadline is December 31 in each year. Contact Miss B. Cox, University House, The University of Lancaster, Bailrigg, Lancaster, LA1 4YW England.

The second International Meeting in Monetary Economics, Banking, and Insurance will be held in Nice, Valbonne, June 6–7, 1985. To present a paper, send two copies before March 9, 1985, to Professors Gaffard and Louberge, CERAM, Sophia Antipolis B.P. 20, 06561 Valbonne, Cedex, France.

The fifth International Symposium on Forecasting, cosponsored by the International Institute of Forecasters and the Faculty of Management of McGill University, will be held in Montreal, June 9–12, 1985. For more information, contact Robert Carbone, Faculty of Management, McGill University, Montreal, PQ, Canada H3A 1G5 (telephone 514+392-4251).

The third annual Multidisciplinary Symposium on American Studies in Africa will be held at the University of Botswana, Gaborone, September 11–14, 1985. Papers are invited relative to the theme, Africa and America: Mutual Perceptions. There is possible funding for participants. Contact Dr. R. F. Morton, University of Botswana, Private Bag 002, Gaborone, Botswana for further information and application forms. The deadline is December 31, 1984 (telephone 51151).

The Russian Research Center, Harvard University, announces fellowships supported by the Andrew W. Mellon Foundation. Applications are being accepted for one annual Senior Mellon Research Fellowship, to be awarded to a scholar at the professorial level; and at least two Mellon Research Fellowships at the assistant professor level. This can include those already engaged in Russian studies, or those in fields of economics, political science, or sociology who are interested in Russian specialization or language fluency. Stipends will be for full support; fellowships are for one year, with possibility of renewal. Applications are due by January 15, 1985. Contact Adam B. Ulam, Director, Russian Research Center, Harvard University, 1737 Cambridge Street, Cambridge, MA 02138 (telephone 617+495-4037).

The GTE Foundation Lectureship Program offers grants of up to $4,000 to accredited colleges and universities to bring in outside lecturers to discuss the broad topic of "Science, Technology, and Human Values." Application forms may be obtained from Richard Schlatter, Director of the Program, Rm 105, 185 College Avenue, Rutgers University, New Brunswick, NJ 08903 (telephone 201+932-7270). The deadline is March 1, 1985.

The Institute of American Cultures Fellowship Program at UCLA offers graduate and postdoctoral fellowships to support study of Afro-Americans, Asian Americans, Chicanos, and American Indians. The graduate fellowship stipend is $5,000 per year plus registration fees and out-of-state tuition if applicable. The postdoctoral fellowship is from $20,000–$23,000 per year (can be awarded for less than a year with stipend adjusted accordingly), and can be used to supplement sabbatical salaries. The deadline for applications is December 31, 1984. For further information and forms, contact the appropriate director in Campbell Hall, UCLA, Los Angeles, CA 90024: Dr. Claudia Mitchell-Kernan, Center for Afro-American Studies; Dr. Lucie Cheng, Asian American Studies Center; Dr. Juan Gómez-Quiñones, Chicano Studies Research Center; Dr. Charlotte Heth, American Indian Studies Center.

The sixth International Meeting of the French Finance Association will be held at INSEAD, Fontainebleau June 13–14, 1985. Send one copy of proposed papers, or indicate other participation, no later than February 1, 1985, to Professor Gabriel Hawawini, INSEAD, Boulevard de Constance, 77305 Fontainebleau, Cedex, France.

Economists who are strongly oriented toward the humanities, who use humanistic methods in their research, and who will be participating in meetings held outside the United States, Mexico, and Canada that are concerned with the humanistic aspects of their discipline are eligible to apply for small travel grants from the American Council of Learned Societies. Financial assistance is limited to air fare between major commercial airports and will not exceed one-half of projected economy-class fare. Social scientists and legal scholars who specialize in the history or philosophy of their disciplines are eligible if the meeting they wish to attend is so oriented. Applicants must hold a Ph.D. degree or its equivalent, and must be citizens or permanent residents of the United States. To be eligible, proposed meetings must be broadly international in sponsorship or participation, or both. The deadlines for application to be received in the ACLS office are for meetings scheduled between July and October, March 1; meetings scheduled between November and February, July 1; meetings scheduled between March and June, November 1. Please request application forms by writing directly to the ACLS (Att: Travel Grant Program), 800 Third Avenue, New York, NY 10022, setting forth the name, dates, place, and sponsorship of the meeting, as well as a brief statement describing the nature of your proposed role in the meeting.

Deaths

Virginia L. Galbraith, professor of economics, Mount Holyoke College, March 24, 1984.

Lionel Robbins, London School of Economics, May 15, 1984.

Retirements

W. D. G. Hunter, professor of economics, McMaster University, June 30, 1984.

J. Richard Powell, professor of economics, California State University-Long Beach, May 31, 1984.

Reuben A. Zubrow, professor of economics emeritus, University of Colorado-Boulder, May 1984.

Foreign Scholars

Willem Buiter, London School of Economics: visiting professor, Yale University, September 1984.

Izzat I. Ghurani, Birzeit University: visiting scholar, Center for Middle Eastern Studies, University of Texas-Austin, 1983–84.

Tapas Majumdar, J. Nehru University: Cornell visiting professor, Swarthmore College, January–June 1985.

Promotions

Neil O. Alper: associate professor of economics, Northwestern University, July 1, 1984.

Robert A. Becker: associate professor of economics, Indiana University-Bloomington, July 1, 1984.

William E. Becker: professor of economics, Indiana University-Bloomington, July 1, 1984.

Susan Schmidt Bies: senior vice president, chief financial officer, First Tennessee National Corporation, May 1984.

Ronald E. Deiter: associate professor of economics, Iowa State University, July 1, 1984.

William M. Edwards: associate professor of economics, Iowa State University, July 1, 1984.

Thomas J. Espenshade: director, Program in Demographic Studies, The Urban Institute, June 1984.

Chuen-mei Fan: associate professor of economics, Colorado State University, July 1, 1984.

Gasper A. Garofalo: associate professor of economics, University of Akron, September 1984.

Gary A. Gigliotti: associate professor of economics, Rutgers University, July 1, 1984.

Barry Ickes: assistant professor of economics, Pennsylvania State University, May 1, 1984.

Gerald D. Jaynes: professor of economics and Afro-American studies, Yale University, July 1984.

Mark R. Killingsworth: associate professor of economics, Rutgers University, July 1, 1984.

Mark Kuperberg: associate professor of economics, Swarthmore College, September 1984.

Devinder M. Malhotra: associate professor of economics, University of Akron, September 1984.

Olivia S. Mitchell: associate professor of economics, Cornell University, May 29, 1984.

Steven C. Myers: associate professor of economics, University of Akron, September 1984.

William S. Peirce: professor of economics, Case Western Reserve University, July 1, 1984.

Kenneth J. Saulter: director, policy and planning, government and public affairs department, Standard Oil Company (Ohio), January 1, 1984.

Roger B. Skurski: professor of economics, University of Notre Dame, 1984–85.

James B. Stewart: associate professor of economics, Pennsylvania State University, July 1, 1984.

Thomas H. Tietenberg: professor of economics, Colby College, September 1, 1984.

Willard E. Witte: associate professor of economics, Indiana University-Bloomington, July 1, 1984.

Administrative Appointments

Donald J. Brown: chairman, economics department, Yale University, July 1984.

Bo Axel Carlsson, Industrial Institute for Economic and Social Research, Sweden: chairman of economics department and professor of industrial economics, Case Western Reserve University, July 1, 1984.

John F. Chizmar: assistant provost, Illinois State University, July 1, 1984.

Charles Craypo, Cornell University: chairman, economics department, University of Notre Dame, fall 1984.

Irving Leveson, Hudson Institute: senior vice president and director of research, Hudson Strategy Group, April 15, 1984.

Lee R. McPheters: director, Bureau of Business and Economic Research, Arizona State University, June 1984.

James Marchand: chairman, department of economics and finance, and professor of economics and finance, University of Mississippi, fall 1984.

Barron A. Peake: chairman, economics department, Golden Gate University, July 1, 1984.

Robert Robertson: chairman, economics department, Mount Holyoke College, July 1984.

F. M. Scherer: chairman and professor of political economy, Swarthmore College, September 1984.

Joseph J. Seneca: chairman, department of economics, Rutgers University, July 1, 1984.

Dennis R. Starleaf: chairman, economics department, Iowa State University, July 1, 1984.

Robert J. Thornton: chairman, economics department, Lehigh University, July 1, 1984.

Appointments

G. J. Anderson: associate professor of economics, McMaster University, July 1984.

M. J. Browning: associate professor of economics, McMaster University, July 1984.

Colleen Callahan, University of North Carolina: assistant professor of economics, Lehigh University, September 1, 1984.

Ching Meei Chang: assistant professor of economics, Rutgers University, July 1, 1984.

N. Edward Coulson, University of Maine: assistant professor of economics, Pennsylvania State University, August 19, 1984.

Dean Croushore, Ohio State University: assistant professor of economics, Pennsylvania State University, August 19, 1984.

Joseph V. Farrell, Massachusetts Institute of Technology: economics, operations research and statistics department, GTE Laboratories Inc., September 1984.

Stefano Fenoaltea: visiting associate professor of economics, Swarthmore College, January–June 1985.

Kwok-Chiu Fung, University of Wisconsin: assistant professor of economics, Mount Holyoke College, September 1984.

Teresa M. Ghilarducci: assistant professor of economics, University of Notre Dame, 1983–84.

Rodney J. Gretlein, Tulane University: assistant professor of economics, Rutgers University, September 1, 1984.

Frank Gunter, John Hopkins University: assistant professor of economics, Lehigh University, August 27, 1984.

Thomas D. Hopkins, U.S. Office of Management and Budget: visiting associate professor, School of Public Affairs, University of Maryland, 1984–85.

Takanobu Ikeda, University of Minnesota: assistant professor of economics, Rutgers University, September 1, 1984.

Renu Kallianpur: visiting assistant professor of economics, Rensselaer Polytechnic Institute, September 1984.

Sun-Young Kim: assistant professor of economics, Illinois State University, August 15, 1984.

Robert W. Kling: assistant professor of economics, Colorado State University, August 20, 1984.

Nicholas Kozlov, University of New Hampshire: assistant professor of economics, Mount Holyoke College, September 1984.

Kathy Langwell, senior economist, Mathematica Policy Research, Inc., July 1983.

Anat Levy, University of California-Los Angeles: assistant professor of economics, Rutgers University, September 1, 1984.

Sharon Long, University of North Carolina: economist, Mathematica Policy Research, Inc., July 1984.

Nelson Lyle, U.S. Department of Health and Human Services: economist, Mathematica Policy Research, Inc., June 1984.

Duncan P. Mann, University of Pennsylvania: assistant professor of economics, Rutgers University, September 1, 1984.

Catherine A. Melfi, University of North Carolina-Chapel Hill: lecturer, economics department, Indiana University-Bloomington, August 1984.

Robert Mendelsohn, University of Michigan: associate professor of forestry and economics, Yale University, July 1984.

Edward M. Miller, Rice University: professor of economics and finance, University of New Orleans, August 1984.

Vivek Moorthy, University of California-Los Angeles: assistant professor of economics, Pennsylvania State University, August 19, 1984.

Ian Novos, University of Pennsylvania: assistant professor of economics, University of Southern California, fall 1984.

Paul Ong: assistant professor of economics, University of California-Santa Cruz, July 1, 1984.

M. J. Osborne: assistant professor of economics, McMaster University, July 1984.

Margaret A. Patterson: assistant professor of economics, California State University-Long Beach, August 27, 1984–May 30, 1985.

Gregory G. Pickett: assistant professor of economics, California State University-Long Beach, August 27, 1984–May 30, 1985.

Herman Quirmbach, Rand Corporation: assistant professor of economics, University of Southern California, fall 1984.

Paul Rabideau, Rutgers University: visiting assistant professor of economics, Swarthmore College, September 1984–June 1985.

Francisco L. Rivera-Batiz, University of Massachusetts: associate professor of economics, Indiana University-Bloomington, August 1984.

Judy Roberts, University of Michigan: economist, Mathematica Policy Research, Inc., October 1984.

Steven M. Rock, Illinois Institute of Technology: visiting associate professor of economics, Rensselaer Polytechnic Institute, September 1984.

Alan J. Rogers, Princeton University: assistant professor of economics, Indiana University-Bloomington, August 1984.

Richard Sabot, IBRD: visiting professor of economics, Williams College, 1984–85.

Raaj K. Sah, Yale University: assistant professor of economics, University of Pennsylvania, July 1984.

William Schulze: professor of economics, University of Colorado-Boulder, September 1984.

Tayyeb Shabbir, University of Pennsylvania: assistant professor of economics, Pennsylvania State University, August 19, 1984.

Matthew Shapiro, Massachusetts Institute of Technology: assistant professor of economics, Yale University, July 1984.

Steven J. Shulman, University of Massachusetts-Amerst: assistant professor of economics, Colorado State University, August 20, 1984.

Padmanabhan Srinagesh, University of Illinois-Chicago: visiting assistant professor of economics, Williams College, fall 1984.

William J. Strang, University of Washington: assistant professor of economics, Washington University-St. Louis, July 1, 1984.

John Strauss, University of Virginia: assistant professor of economics, Yale University, July 1984.

Larry Taylor, University of North Carolina: assistant professor of economics, Lehigh University, September 1, 1984.

David D. VanHoose, University of North Carolina-Chapel Hill: assistant professor of economics, Indiana University-Bloomington, August 1984.

Mark Walbert: assistant professor of economics, Illinois State University, August 15, 1984.

Donald Waldman: assistant professor of economics, University of Colorado-Boulder, September 1984.

James M. Walker, University of Arizona: assistant professor of economics, Indiana University-Bloomington, August 1984.

David Weiman, Swarthmore College: assistant professor of economics, Yale University, July 1984.

Arthur L. Welsh, Joint Council on Economic Education: visiting professor of economics, Indiana University-Bloomington, August 1984.

Stephen Younger, Stanford University: assistant professor of economics, Williams College, July 1984.

Leaves for Special Appointments

Donald C. Cox, Washington University-St. Louis: Hoover Institution Fellow, 1984–85.

Michael Nelson, Illinois State University: Marquette University, August 1984–May 1985.

David Shapiro, Pennsylvania State University: USAID, Zaire, July 1, 1984–June 30, 1985.

Joe A. Stone, University of Oregon: Council of Economic Advisors, 1984–85.

Gottfried E. Voelker, Fachhochschule Osnabrueck: Adviser to the Minister, Ministry of Planning and Economic Affairs Monrovia, Liberia, Africa, August 1, 1984–July 30, 1985.

Michael J. Wasylenko, Pennsylvania State University: Syracuse University, July 1, 1984–June 30, 1985.

Michael Wyzan, Illinois State University: Fulbright Fellow, Bulgaria, January 1985–May 1985.

Resignations

Richard Coe, University of Notre Dame: New College of University of South Florida, fall 1984–85.

Samar Datta, University of Southern California: Accro-Economic Research Center, India, fall 1984.

Carlos Diaz-Alejandro, Yale University: Columbia University, June 30, 1984.

Jonathan Eaton, Yale University: University of Virginia, June 30, 1984.

David E. Harrington, University of Notre Dame: Pomona College, fall 1984.

Laurence G. Kantor, Lehigh University: Federal Reserve Bank, August 31, 1984.

Laurence J. Kotlikoff, Yale University: Boston University, June 30, 1984.

Kathy J. Krynski, University of Notre Dame: Pomona College, fall 1984.

Christopher Lingle, Miami University, February 1984.

Anu Luthur, University of California-Santa Cruz, July 1, 1984.

Yash Mehra, Pennsylvania State University: Federal Reserve Bank of Richmond, July 1, 1984.

Robert A. Moffitt, Rutgers University: Brown University, July 1, 1984.

Hugh T. Patrick, Yale University: Columbia University, June 30, 1984.

NOTE TO DEPARTMENTAL SECRETARIES AND EXECUTIVE OFFICERS

When sending information to the *Review* for inclusion in the Notes Section, please use the following style:

A. Please use the following categories (please—do not send public relation releases):

1—Deaths
2—Retirements
3—Foreign Scholars (visiting the USA or Canada)
4—Promotions
5—Administrative Appointments
6—New Appointments
7—Leaves for Special Appointments (NOT Sabbaticals)
8—Resignations
9—Miscellaneous

B. Please give the name of the individual (SMITH, Jane W.), her present place of employment or enrollment: her new title (if any), new institution and the date at which the change will occur.

C. Type each item on a separate 3×5 card.

D. The closing dates for each issue are as follows: *March*, October 15; *June*, January 15; *September*, April 15; *December*, July 15.

All items and information should be sent to the Assistant Production Editor, *American Economic Review*, Room 8279, Bunche Hall, University of California, Los Angeles, CA 90024.

EIGHTY-FIRST LIST OF DOCTORAL DISSERTATIONS IN POLITICAL ECONOMY IN AMERICAN UNIVERSITIES AND COLLEGES

The present list specifies doctoral degrees conferred during the year terminating June 1984.

General Economics; including Economic Theory, History of Thought, Methodology, Economic History, and Economic Systems

DILIP ABREU, Ph.D. Princeton 1983. Repeated games with discounting: A general theory and an application to oligopoly.

MARGARET C. AGNEW, Ph.D. Massachusetts Institute of Technology 1983. The role of the medium of exchange in a general disequilibrium model.

JACK L. AMARIGLIO, Ph.D. Massachusetts 1984. Economic history and the theory of primitive society.

DAVID A. ASCHAUER, Ph.D. Rochester 1984. Essays in macroeconomics.

ROBERT J. BATEMARCO, Ph.D. Georgetown 1984. Studies in the Austrian theory of the cycle.

MICHAEL R. BAYE, Ph.D. Purdue 1983. Three essays on price indices, price dispersion, and consumers' welfare.

TERRENCE M. BELTON, Ph.D. Michigan 1983. Entry and entry deterrence under uncertainty: Three essays on firm behavior under asymmetric information.

JOHN H. BIGGS, Ph.D. Washington-St. Louis 1983. The demand for life insurance.

SEVERIN J. BORENSTEIN, Ph.D. Massachusetts Institute of Technology 1983. Price discrimination in free-entry markets.

RICHARD M. CANTOR, Ph.D. Johns Hopkins 1984. Three essays on implicit contracts and macroeconomics.

PATRICIA K. CENTRELL, Ph.D. Southern Methodist 1983. Interregional exchange rates in the antebellum United States.

ANDREW S. CAPLIN, Ph.D. Yale 1983. Aggregation of (S, s) inventory policies.

JOHN P. CASKEY, Ph.D. Stanford 1984. Modelling the formation of price expectations.

IMAD A. CHATILA, Ph.D. Southern California 1983. Discount rates for public investments under uncertainty.

MARIE-THÉRÈSE CHICHA-PONTBRIAND, Ph.D. McGill 1983. An extended approach to recording economic surplus: Conceptual, quantitative, applied issues.

YOON JE CHO, Ph.D. Stanford 1984. On the liberalization of the financial system and efficiency of capital allocation under uncertainty.

CHIN-FU CHUANG, Ph.D. State University of New York-Buffalo 1984. A Nash-bargaining model of household demand.

MUN PAE CHUN, Ph.D. George Washington 1984. Price determination and inflation in an open economy.

RAMON L. CLARETE, Ph.D. Hawaii-Manoa 1984. The costs and consequences of trade distortions in a small open economy: A general equilibrium model of the Philippines.

JEFFREY L. COLES, Ph.D. Stanford 1983. Two essays in economic theory.

XAVIER COMAS, Ph.D. Johns Hopkins 1984. The allocation of scarce resources to life-saving programs with an appendix on the economic value of human life.

BARRY V. COZIER, Ph.D. Western Ontario 1984. A rational expectations equilibrium model of a small, specialized economy.

ASHIS K. DEV, Ph.D. State University of New York-Stony Brook 1983. A general equilibrium analysis of the time structures of saving, investment, and financial decisions.

BEHZAD TABATABAI DIBA, Ph.D. Brown 1984. Essays on economic equilibria. Part I: Majority voting equilibria in the public sector; Part II: Rational asset price bubbles: Theory and evidence.

EUGENE DYKEMA, Ph.D. Notre Dame 1984. Methodology and scope in Gunnar Myrdal's economy.

MARY E. ELLIS, Ph.D. South Carolina 1983. The comparison and evaluation of risk measures as proxies for investor risk.

TERRY ELROD, Ph.D. Columbia 1984. Inferring a market map from observed choice behavior.

UGO FASANO FILHO, Ph.D. Illinois-Urbana 1984. Currency substitution, interest rate, and exchange rate.

DAVID FELDMAN, Ph.D. Northwestern 1984. A theory of asset prices and the term structure of interest rates in a partially observable economy.

JAMES T. FERGUS, JR., Ph.D. Columbia 1984. Substitution between direct and indirect automobile financing: Sources, extent, and policy implications.

MARIO FORTUNA, Ph.D. Boston College 1983. Separability in the model of household demand for financial assets under uncertainty.

ODED GALOR, Ph.D. Columbia 1984. Three essays in economic theory.

ISRAEL GERMAN, Ph.D. Purdue 1983. A disequilibrium macro model and catastrophe theory: The case of an oil shock.

DAVID GLEICHER, Ph.D. Columbia 1984. Critical studies in the origins of classical value theory.

PAUL A. GOLDBERG, Ph.D. Columbia 1984. Essays on industrial product quality as reflected in the literature of Soviet economists.

DEBORAH HAAS-WILSON, Ph.D. California-Berkeley 1983. Asymmetric information, regulation, and qualify-adjustment prices: The case of optometry.

JAMES HERREMAN, Ph.D. Indiana-Bloomington 1983. A bilateral expectations policy analysis model.

JOSEPH A. HERRIGES, Ph.D. Wisconsin-Madison 1983. Issues in the measurement of consumer response to time-of-use electricity pricing.

GREGORY W. HUFFMAN, Ph.D. Minnesota 1983. An overlapping generations model of asset pricing.

FREDERIK H. HUIZINGA, Ph.D. Massachusetts Institute of Technology 1984. Heterogeneous labor, relative wages, and macroeconomics.

BARRY W. ICKES, Ph.D. California-Berkeley 1984. A macroeconomic theory of centrally planned economics.

TAKANOBU IKEDA, Ph.D. Minnesota 1984. On the highest degree of smoothness of outcome functions compatible with non-Walrasian performance.

DONALD R. JONES, Ph.D. Massachusetts Institute of Technology 1984. Contractual solutions to problems of risk sharing and moral hazard.

MOONSOO KANG, Ph.D. Minnesota 1984. Resource supply shocks and the interest rate.

MYUNG H. KANG, Ph.D. State University of New York-Albany 1984. A study in macroeconomic causality.

IMRE KARAFIATH, Ph.D. Texas A&M 1983. An analytic survey of the development of business cycle theory.

STEPHEN B. KEALHOFER, Ph.D. Princeton 1983. Liquidity, information, and production.

ALEXANDER S. KELSO, JR., Ph.D. Massachusetts Institute of Technology 1984. Essays in economic theory.

WAGAR M. KHAN, Ph.D. Boston 1984. Towards an interest-free Islamic economic system: A theoretical analysis of prohibiting debt financing.

STEPHEN R. KING, Ph.D. Northwestern 1984. Macroeconomic activity and the real rate of interest.

DANIEL J. KOVENOCK, Ph.D. Wisconsin-Madison 1983. Essays in economic theory.

DAVID A. LEREAH, Ph.D. Virginia 1983. Information problems and regulation in insurance markets.

JASPER LESAGE, Ph.D. Toronto 1984. The European-Nootkan maritime fur trade in the late eighteenth century.

JOSEPH W. MCANNENY, Ph.D. George Washington 1984. Spatial competition.

KENNETH F. MCCUE, Ph.D. California Institute of Technology 1984. The structure of individual decisions in American elections: The influence of relevant alternatives.

W. BENTLEY MACLEOD, Ph.D. British Columbia 1984. Perspectives on oligopoly theory.

SCOTT MAGRUDER, Ph.D. Southern Illinois-Carbondale 1984. A simulation of a hypothetical general equilibrium economy with micro foundations.

KEVIN J. MALONEY, Ph.D. Washington-St. Louis 1983. Inflation, taxes, and the productive capacity of the U.S. economy.

STEPHEN V. MARKS, Ph.D. Princeton 1984. Market power and imperfect information: The international crude oil market.

DONALD J. MEYER, Ph.D. Texas A&M 1983. The effect of competition in sealed-bid auctions: Certainty and uncertainty.

WAYNE MILLER, Ph.D. Wisconsin-Madison 1983. Material well-being: An exploratory study of the relationship between resource endowments, institutional arrangements and material well-being.

JANET R. MINIFIE, Ph.D. South Carolina 1983. An experimental investigation of material requirements planning system nervousness.

CHARLES S. MORRIS, Ph.D. California-Los Angeles 1983. Cyclical productivity in the United States: A reconciliation of theory and evidence.

STEVE D. MULLINS, Ph.D. Oklahoma State 1983. A prestige maximization model of institutions of higher learning.

HOWARD F. NAISH, Ph.D. Southern California 1983. Imperfect competition, price adjustment costs, and the Phillips curve.

LYLE M. NELSON, Ph.D. Wisconsin-Madison 1984. The welfare effects of a shift from income to commodity taxation.

YOSHIMASA NOMURA, Ph.D. Johns Hopkins 1984. Approximate equilibria with infinitely many commodities.

ALFONSO NOVALES, Ph.D. Minnesota 1984. A stochastic life cycle theory of consumption.

TATSUYA OHMORI, Ph.D. Southern California 1983. The Japanese economy: A modified form of contemporary managed capitalism.

JOHN F. OLSON, Ph.D. Rochester 1984. The occupational structure of plantation slave labor in the late antebellum era.

JAMES K. O'TOOLE, Ph.D. Virginia Polytechnic Institute 1984. A model of control agency behavior.

SPENCER J. PACK, Ph.D. New Hampshire 1983. The theoretical genesis of capital using a commodity theory of value.

DAVID G. PEARCE, Ph.D. Princeton 1983. Two essays on the theory of strategic behavior.

ERIC B. RASMUSEN, Ph.D. Massachusetts Institute of Technology 1984. Essays in dynamic nonprice competition.

ALISTAIR R. ROBERTSON, Ph.D. Queen's 1984. Price setting and quantity adjustment in atomistic markets.

AILSA A. RÖELL, Ph.D. Johns Hopkins 1984. Essays in the theory of comparative risk aversion: With special reference to the case of state-dependent preferences.

ANTONIN RUSEK, Ph.D. State University of New York-

Stony Brook 1984. Production functions which incorporate managerial inputs.

MOHAMMAD R. SAFARZADEH, Ph.D. Claremont 1983. Money growth and the behavior of industrial prices.

GEOFFREY F. SANDERS, Ph.D. Colorado-Boulder 1983. A varying parameter approach to interest rate determination in a term structure model.

WILFREDO SANTIAGO-VALIENTE, Ph.D. Columbia 1984. Three essays in the early twentieth-century American theory of money, capital, and banking.

STEPHEN R. SELINGER, Ph.D. California Institute of Technology 1984. Three essays in theoretical, applied, and normative economics.

YOCHANAN SCHACHMUROVE, Ph.D. Minnesota 1983. A rational expectations equilibrium model of capital accumulation under uncertainty.

MATTHEW D. SHAPIRO, Ph.D. Massachusetts Institute of Technology 1984. The dynamic demand for capital and labor and other essays in macroeconomics.

STEPHEN C. SHEPPARD, Ph.D. Washington-St. Louis 1984. Existence of equilibrium in spatial economics with a continuum of consumers.

ALAN SLEEMAN, Ph.D. Simon Fraser 1983. Essays on the United Kingdom Phillips curve.

CHERYL I. SMITH, Ph.D. Yale 1984. Theories of fluctuations in the monetary economy: A study of the concepts of money, capital, and investment in the theories of Wicksell, Hayek, and Keynes.

WILLIAM G. STANFORD, Ph.D. Northwestern 1984. Conjectural variations and oligopoly supergames.

VALARIE Y. SUSLOW, Ph.D. Stanford 1983. Monopoly pricing with intertemporal complications.

PAUL A. SWANSON, Ph.D. Columbia 1984. Essays on value theory: Fixed capital and socially necessary labor reexamined.

HAMID R. TASLIMI, Ph.D. State University of New York-Stony Brook 1983. Macroeconomic policy analysis for an open economy with quantity constraints.

LEWIS A. TAYLOR III, D.B.A. Indiana 1984. Strategic alternative and goal level effects on decision quality and goal commitment.

LUCILE S. THOMPSON, Ph.D. California-San Diego 1983. Structural changes in the aggregate relation between consumption expenditures and income in the United States, 1929–81.

DAVID KWAI CHE TSE, Ph.D. California-Berkeley 1984. A model of consumer post-choice processes.

GREGORY J. TULLY, Ph.D. California-Berkeley 1983. Characteristic responses of information structures to incentive variations.

F. JAVIER VIVES, Ph.D. California-Berkeley 1983. Cournot and Bertrand competition in product differentiation models.

KARL VOLTAIRE, Ph.D. Chicago 1983. Two essays in stochastic capital theory.

KENJI YAMAMOTO, Ph.D. State University of New York-Buffalo 1984. Pareto-efficient allocations and competitive equilibrium in overlapping-generations models.

NICHOLAS C. YANNELIS, Ph.D. Rochester 1983. Solution concepts in economic theory.

ROBERT M. YARBROUGH, Ph.D. Washington 1983. Three essays in search and production theory.

MENG-HUA YE, Ph.D. Wisconsin-Madison 1984. Applications of Brownian motion to economic models of optimal stopping.

MINE K. YÜCEL, Ph.D. Rice 1984. An optimal control approach to the taxation of exhaustible resources.

THOMAS L. WALKOM, Ph.D. Toronto 1983. The daily newspaper industry in Ontario's developing capitalist economy: Toronto and Ottawa, 1871–1911.

SHINICHI WATANABE, Ph.D. Minnesota 1983. Dynamic function of the price mechanism.

ERNST J. WEBER, Ph.D. Rochester 1984. The Great Depression in Switzerland, Sweden and the United States: The money hypothesis vs. the spending hypothesis.

MICHAEL D. WHINSTON, Ph.D. Massachusetts Institute of Technology 1984. Influence, common agency, and strategic behavior.

GILLIAN WOGIN, Ph.D. Carleton 1984. The wealth-maximizing behavior of the Canadian Pacific Railway: Lands, freight rates, and the Crow's Nest Pass Agreement.

MICHAEL D. WOODFORD, Ph.D. Massachusetts Institute of Technology 1983. Essays in intertemporal economics.

DEMETRIUS C. YANNELIS, Ph.D. Queen's 1983. Essays on disequilibrium dynamics.

IGNACIO ZUBIRI, Ph.D. Princeton 1983. Three essays on the theory of income distribution.

Economic Growth and Development; including Economic Planning Theory and Policy; Economic Fluctuation and Forecasting

ABDULFATTAH A. ABOHOBIEL, Ph.D. Indiana-Bloomington 1983. An econometric model for the Libyan economy, 1962–77.

SALAHEDDIN S. ABOSEDRA, Ph.D. Colorado-Boulder 1984. A simulation study of the impact of world oil prices on the development of the Libyan economy.

SADIQ AHMED, Ph.D. Boston 1984. Inflation in Bangladesh: Causes and consequences.

Y. AL-EBRAHEEM, Ph.D. Claremont 1984. The optimal rate of oil production and economic development: The case of Kuwait.

PAULO DE TARSO A. ANDRE, Ph.D. Illinois-Urbana 1983. Consumer choice among durables: A random utility model and an application to Brazilian households.

CHARUMA ASHAKUL, Ph.D. Colorado-Boulder 1983. In-

come distribution in Thailand: Analysis with policy implications.

TERRA ASHAKUL, Ph.D. Colorado-Boulder 1983. Estimates of an acreage response model of five major annual crops in Thailand, 1967–81.

JOSEPH BAAH-DWOMOH, Ph.D. Western Ontario 1984. Sources of growth and change in the economy of Ghana, 1900–65.

YOUSEF M. BADI, Ph.D. Nebraska-Lincoln 1982. Regional growth impact of Libyan development budget expenditures.

THEOLOGOS H. BONITSIS, Ph.D. City-New York 1984. Economic development and external flows: The South American experience.

ALEXANDER BOWEN, Ph.D. Massachusetts Institute of Technology 1984. Adjusting to city life: Essays on the transition in East Africa.

LOREN BRANDT, JR., Ph.D. Illinois-Urbana 1983. Population growth and agricultural change in central and eastern China, 1898–1936.

RICKY L. CHANEY, Ph.D. Illinois-Urbana 1984. Emigrant remittance and economic and demographic disparities across regions in Portugal.

AJAY CHHIBBER, Ph.D. Stanford 1984. The sources of inflation and anti-inflationary policy in India.

IRWIN L. COLLIER, JR., Ph.D. Massachusetts Institute of Technology 1984. Connections, effective purchasing power, and real product in the German Democratic Republic.

PATRICK J. CONWAY, Ph.D. Princeton 1984. The interaction of government and private-sector responses to an exogenous economic shock: Turkey after 1973.

KEITH W. CRANE, Ph.D. Indiana-Bloomington 1983. A comparison of foreign trade decision making in Poland and Hungary.

MARIA A. CRUZ-SACO OYAGUE, Ph.D. Pittsburgh 1983. Estimation of a money-output model for the Peruvian economy.

CHRISTINE M. CUMMING, Ph.D. Minnesota 1983. An international comparison of income distribution.

HARENDRA DEY, Ph.D. Boston 1984. Changes in real wages: The effect of growth, the consequences of income distribution.

ABDUNASSER DUELLA, Ph.D. Cincinnati 1983. An econometric model of Iraq, 1960–76.

AKPAN HOGAN EKPO, Ph.D. Pittsburgh 1983. An economic study of the food crop economy of a petroleum-based economy: The Nigerian case.

FAKRY A. ELFIKI, Ph.D. Clark 1984. Foreign economic assistance and the Egyptian economy.

E. PHILIP ENGLISH, Ph.D. Toronto 1983. The role of tourism in the economic development of Senegal.

THOMAS EPONOU, Ph.D. Michigan State 1983. Farm level analysis of rice production systems in northwestern Ivory Coast.

ROBERT A. FELDMAN, Ph.D. Massachusetts Institute of Technology 1984. Japanese financial markets: Innovation and evolution.

STEVEN C. FRANZEL, Ph.D. Michigan State 1983. Planning an adaptive production research program for small farmers: A case study of farming systems research in Kirinyage District, Kenya.

RICARDO FRICKE-URQUIOLA, Ph.D. Texas-Austin 1983. Spatial development in Mexico: Regional inequalities, decentralization of the manufacturing sector, and regional policies.

MOHAMMED H. GHULOUM, Ph.D. Claremont 1984. A model of the monetary sector of Kuwait.

GUSTAVO P. D. M. G. GOMES, Ph.D. Illinois-Urbana 1984. The roots of state intervention in the contemporary Brazilian economy.

STEVEN L. GREEN, Ph.D. Brown 1984. Excess inflation in three models with rational monetary policy.

CARL H. GROTH, JR., Ph.D. George Washington 1984. Alliance theory: The role of private vs. public defense goods.

EDWIN GUITERREZ, Ph.D. Oklahoma State 1984. A policy analysis of industrialization and foreign trade for Costa Rica.

NIMAL G. GUNATILLEKE, Ph.D. Michigan State 1983. An economic analysis of housing and home-based business in Columbo, Sri Lanka.

STEVEN HAGGBLADE, Ph.D. Michigan State 1984. Sorghum beer in Botswana: The impact of factory brews on a cottage industry.

GREGORY HAMILTON, Ph.D. Oklahoma State 1983. An analysis of cumulative urban and region growth.

HAMZAID, Ph.D. Washington State 1984. An econometric model of the demand for money in an open developing economy: The Malaysian case.

ANN HELWEGE, Ph.D. State University of New York-Buffalo 1984. Constraints on the economic dependence of developing countries.

CARLOS D. HEYMANN, Ph.D. California-Los Angeles 1983. A study in economic instability: The case of Argentina.

JOHN HITCHINGS, Ph.D. Stanford 1982. Agricultural determinants of nutritional status among Kenyan children.

MOHAMMAD H. HODA, Ph.D. McMaster 1984. Macroeconometric and input-output model of Iran.

KYTTACK HONG, Ph.D. Stanford 1983. Monetary policies in semi-industrial countries: Theory and evidence from Korea and Taiwan.

YOO SOO HONG, Ph.D. Northwestern 1983. Intra-industrial planning and control with an application to the Korean cement industry.

OH SEOK HYUN, Ph.D. Pennsylvania 1984. A macroeconometric model of Korea: Simulation experiments with a large scale model for a developing country.

SALISU A. INGAWA, Ph.D. Michigan State 1983. Market-

ing of food crops and inputs: The case of Funtua agricultural development project in Kaduna State, Nigeria.

ASHFAQ M. ISHAQ, Ph.D. George Washington 1984. Measurement of firm performance and analysis of its determinants: A study of small firms in Colombia.

PREECHA JARUNGIDANAN, Ph.D. Missouri-Columbia 1983. Test of the appropriate technology model for economic development.

ALIMAMI M. KARGBO, Ph.D. Michigan State 1983. An economic analysis of rice production systems and production organization of rice farmers in the Gambia.

S. A. M. AMIN KIANIAN, Ph.D. Colorado-Boulder 1983. Estimation and projection of the demand for refined petroleum products in Iran.

IL-CHUNG KIM, Ph.D. Kansas 1984. Environmental policy and public finance in a neoclassical model of optimal growth.

JEFFREY G. KITIGAN, Ph.D. Tufts 1984. Political stability and economic development in Malaysia.

BARRY A. KRISSOFF, Ph.D. Virginia 1983. An empirical analysis of the impact of external resource transfers on the development process.

KATHIE L. KRUMM, Ph.D. Stanford 1983. Investment during an attempted liberalization: Argentina, 1976–81.

YUZO KUMASAKA, Ph.D. Pennsylvania 1984. A comparison of the slowdown in productivity growth after the two oil crises: Japan and the United States.

BRUCE L. KUTNICK, Ph.D. Massachusetts Institute of Technology 1984. The determinants of industrial research and development in a developing country: The case of India.

LAISSE LAHOUANI, Ph.D. Cincinnati 1984. Macroeconomic policy and stabilization program in a small open economy: An empirical analysis of the foreign trade pattern of the Ivory Coast, 1963–80.

ENRIQUE J. LAMAS, Ph.D. George Washington 1984. The impact of military installations on local economies.

NGUYURU H. I. LIPUMBA, Ph.D. Stanford 1983. Foreign exchange and economic development in Tanzania.

JIMMY LOVE, Ph.D. Oklahoma State 1983. An analysis of economic growth in the Dominican Republic, 1966–75: The effects on the distribution of income and regional employment.

JAMES H. MCCLURE, JR., Ph.D. Vanderbilt 1983. Fiscal policy reaction functions for the group of ten: A cross-country comparison of demand policy and the inflation-unemployment dilemma.

JUDITH R. MCKINNEY, Ph.D. Indiana-Bloomington 1983. Economic exploitation of nonrenewable resources in a centrally planned economy: An investigation of the use of natural resources in the U.S.S.R.

MICHAEL P. MCLINDON, Ph.D. Tufts 1984. Cocoa in Cameroon: Policy and economics.

SHAHRAM MANOUCHEHRI, Ph.D. Alberta 1982. Inflation in less developed countries: The case of Iran, 1959–77.

TAHA MANSOUR, Ph.D. Cincinnati 1983. International labor migration: The case of Jordan.

MICHAEL MASSOURAKIS, Ph.D. City-New York 1983. Inflation and stabilization in a small open economy: The case of Greece, 1960–80.

ABDUL M. MAT SALLEH, Ph.D. Hawaii-Manoa 1983. Effects of socioeconomic changes and family planning of fertility in peninsula Malaysia.

VINCENT MENSAH, Ph.D. Wayne State 1984. The impact of expanding financial institutions on investment and real capital formation: The case study of Ghana.

KENT W. MIKKELSEN, Ph.D. Yale 1984. Inventive activity in Philippine industry.

FAIZ MOHAMMAD, Ph.D. Simon Fraser 1983. An analysis of structure and performance of agricultural markets in Pakistan.

YASMEEN N. MOHIUDDIN, Ph.D. Vanderbilt 1983. An investigation into the determinants of farm employment in Pakistan.

LEONARD M. MOITE, Ph.D. California-Los Angeles 1984. Performance of the modern private sectors of the Kenyan economy.

FELIPE G. MORANDE, Ph.D. Minnesota 1983. Retail and wholesale prices in Chile: International commodity arbitrage and the "Law of One Price."

AHMAD H. MUSTAFA, Ph.D. Texas-Austin 1983. The distribution aspects of special drawing rights and the interests of developing countries in a demand for international reserves based distribution.

MANEGA NDULO, Ph.D. Michigan State 1983. An analysis of a low-income housing market in Zambia.

M. KHAN NIAZI, Ph.D. Boston 1984. Economic growth in South Asia.

GLADSON NWANNA, Ph.D. Fordham 1984. Impact of foreign direct investment in Nigeria, 1960–80.

GUMERSINDO OLIVEROS, Ph.D. Minnesota 1984. A quantitative analysis of, the general equilibrium effects associated with tariff removal in Spain.

ARIS OTHMAN, Ph.D. Boston 1984. Growth, equality, and poverty in Malaysia, 1957–80.

MARY C. OTT, Ph.D. California-Davis 1984. The role of the private sector as a provider of industrial training: The case of the Honduran furniture industry.

ISMAEL S. OUEDRAOGO, Ph.D. Michigan State 1984. A socioeconomic analysis of farmers' food grain marketing linkages and behavior in eastern Upper Volta.

EVA A. PAUS, Ph.D. Pittsburgh 1983. Stagnation between export promotion and a protected internal market: The Colombian textile industry.

DOUGLAS H. PERRY, Ph.D. Stanford 1982. A model of the Pakistan dairy sector: Interaction of government policy and technology.

HASSAN PORDELI, Ph.D. Nebraska-Lincoln 1983. To-

ward more effective economic planning in developing nations: An application of the Delphi approach.

BETTY J. PUNNETT, Ph.D. New York 1984. Goal setting and need for achievement: The relationship to motivation in the less developed countries.

GHULAM QADIR, Ph.D. Boston 1984. Determinants of personal earnings in a dual economy and transmission of economic status from generation to generation: An empirical study of Pakistan.

SABIT RAHIM, Ph.D. State University of New York-Binghamton 1984. Benefits to West Pakistan from the economic union with East Pakistan.

MIGUEL D. RAMIREZ, Ph.D. Illinois-Urbana 1984. Development banking in Mexico.

IDRIS RASHIDI, Ph.D. Boston 1984. Policy simulation model of the Tanzanian economy.

MEHDI B. RAZAVI, Ph.D. Nebraska-Lincoln 1983. Modeling an Islamic economic system: An interaction delivery matrix and Boolean Digraph approach.

PEDRO RIVERA, Ph.D. State University of New York-Binghamton 1984. Transfer payments and the retirement and precautionary motives for saving: An evaluation of the impact of U.S. transfers on personal saving in Puerto Rico.

FREDERICK C. ROCHE, Ph.D. Stanford 1983. Cassava production systems on Java and Madura.

DAVID F. RUCCIO, Ph.D. Massachusetts 1984. Optimal planning theory and theories of socialist planning.

HOSSEIN SARJEHPEYMA, Ph.D. Oklahoma 1984. The economic impact of internal migration: Iran as a special case.

AYUBA J. SARKI, Ph.D. Georgia 1984. Domestic credit expansion and the balance of payments problems: Nigeria, 1960–78.

JOSEPH J. SEMBOJA, Ph.D. Illinois-Urbana 1983. Government policy and equity in Tanzania.

GHOLOMALI SHARZEHI, Ph.D. Kansas 1984. Optimal public investment for boomtowns: A welfare economic analysis.

SIAMACK SHOJAI, Ph.D. Fordham 1984. A macroeconomic approach to oil production in OPEC countries.

TERRY SICULAR, Ph.D. Yale 1983. Market restrictions in China.

ANUP K. SINHA, Ph.D. Southern California 1983. Essays on some Lewisian themes of economic development.

DAVID SIVAN, Ph.D. Southern Illinois-Carbondale 1983. The inflationary experience of a developing economy: A monetary analysis of the Israeli economy, 1954–79.

MOHAMMAD SOHRAB-SUDDIN, Ph.D. Vanderbilt 1984. Money, prices, and balance of payments: Studies of selected south Asian countries.

CHARLES STRUCKMEYER, Ph.D. Yale 1983. Capital, energy, and economic growth: A vintage approach.

RICHARD J. SULLIVAN, JR., Ph.D. Illinois-Urbana 1983. English agriculture, 1500–1850: A case study of long-run technological change.

LAURA TANDY, Ph.D. Rutgers 1984. The export side of the senile/injured industry arguments: A case study of Bangladesh jute.

SUVIT THANIYAVARN, Ph.D. Illinois-Urbana 1984. Monetary policy in Thailand: A case study of the monetary authorities' actual behavior, 1960–61.

KARL THEODORE, Ph.D. Boston 1984. Government influence and the enclave economy of Trinidad and Tobago.

STEVEN H. WHITING, Ph.D. Michigan 1983. Essays on the structure of the Egyptian financial system, 1961–81.

ENOCK E. M. WIKETYE, Ph.D. Washington State 1984. The role of financial intermediaries in economic growth in LDCs: The case of eleven African countries.

GHAZI YOUSSEF, Ph.D. Wayne State 1983. Lomè: A closer look at an experimental trade association agreement.

ABDOLHAMID ZANGENEHZADEH, Ph.D. Missouri-Columbia 1984. Stabilization policies in less developed countries: A case study of prerevolutionary Iran.

Economic Statistics; including Econometric Methods, Economic and Social Accounting

CHARLES G. BERESKIN, Ph.D. Missouri-Columbia 1983. Translog cost functions for railroads.

CARLO-ANDREA BOLLINO, Ph.D. Pennsylvania 1983. Three essays in empirical demand analysis.

JAMES A. CHALFANT, Ph.D. North Carolina State 1983. Choosing among flexible functional forms: An application of the generalized Box-Cox and Fourier flexible forms to U.S. agriculture.

ANUP K. CHANDA, Ph.D. Florida 1984. Econometrics of disequilibrium under rational expectations.

CHUN KYUNG CHANG, Ph.D. Oklahoma 1984. An econometric model of monthly peak load: Case study for an electric utility system.

JAIN-SHING ADOLPH CHEN, Ph.D. Missouri-Columbia 1983. Simultaneous equations models with qualitative dependent variables: A food stamp program participation and food cost analysis.

DAYAN DAWOOD, Ph.D. Hawaii-Manoa 1983. Export earnings instability in Indonesia: An intertemporal analysis of the postwar statistical record.

ALEX DONTOH, Ph.D. New York 1984. Public information structures: An analysis of stewardship function.

GALIP FEYZIOGLU, Ph.D. Ohio State 1983. The analysis and assessment of time variant linear trends in annual economic data series with an application to energy forecasting for the state of Ohio.

LUKE M. FROEB, Ph.D. Wisconsin-Madison 1983. The measurement of feedback between bivariate time-series of profits and concentration.

BARRY GALEF, Ph.D. Yale 1983. The diffusion of analytical instruments.

ANDRE GASCON, Ph.D. California-Berkeley 1984. Multi-item scheduling on a single machine.

JUAN J. GONZALEZ, Ph.D. South Carolina 1983. An interactive approach for solving a multiple objective integer programming problem.

ALI S. HADI, Ph.D. New York 1984. Effects of simultaneous deletion of a variable and an observation on a linear regression equation.

JAMES P. HART, JR., Ph.D. California-Berkeley 1984. A Lagrangian relaxation and branch-and-bound procedure for solving problems with work-load balancing constraints.

SEYED A. A. JAZAYERI, Ph.D. Southern California 1984. A new econometric approach to modelling the peak load pricing policies.

MAX E. JERRELL, Ph.D. Arizona State 1984. An econometric model of Arizona.

WILLIAM H. JOYCE, Ph.D. New York 1984. Reinvestment value determination of price: A technique to estimate price for an industrial commodity with positive volume growth rate.

SUNG WHAN JU, Ph.D. State University of New York-Albany 1984. An econometric model of congressmens' voting behavior with sample selectivity.

SUK KANK, Ph.D. Wisconsin-Madison 1983. Estimation of an almost ideal demand system from panel data.

CHIH-HWA DUKE KAO, Ph.D. State University of New York-Stony Brook 1983. Essays on the estimation of heteroscedastic regression models.

CHING-WUU K. KENG, Ph.D. Toronto 1983. Forecasting methodology and applications.

AKBAR KOMIJANI, Ph.D. Wisconsin-Milwaukee 1983. Analysis of the modified money demand function and its stability through alternative econometric techniques.

TENPAO LEE, Ph.D. Iowa State 1984. Transportation cost estimation and its implications.

JOHN P. LINK, Ph.D. Northwestern 1984. The existence of the maximum-likelihood estimate in log-linear probability models.

CORNELIS A. LOS, Ph.D. Columbia 1983. Econometrics of models with evolutionary parameter structures.

BAKHTIAR MOAZZAMI, Ph.D. Alberta 1984. A Monte Carlo study of the small sample properties of the simultaneous equation estimators in the presence of autocorrelated disturbances.

FARID NOVIN, Ph.D. Queen's 1983. A Bayesian choice theoretic approach to the dynamic adjustment of prices: A study of price behavior of Canadian manufacturing industries.

STAVROS C. PERISTIANI, Ph.D. Columbia 1984. The general linear structural econometric model in the presence of limited dependent variables.

PAUL J. PIEPER, Ph.D. Northwestern 1984. The measurement of real investment in structures and the construction productivity decline.

ELIZABETH L. RANKIN, Ph.D. Virginia Polytechnic Institute 1984. Aliasing and the estimation of distributive lag models.

PETER E. ROSSI, Ph.D. Chicago 1984. Specification and analysis of econometric production models.

SUNIL K. SAPRA, Ph.D. Columbia 1983. Three essays in dynamic econometrics.

WILLIAM C. SAWYER, Ph.D. New York 1983. Correlation coefficient estimation and index models.

WILLIAM R. SINGLETON, Ph.D. Hawaii-Manoa 1983. An empirical study of the effects of taxation on investment expenditures by selective firms in the forest products industry.

DANIEL J. SLOTTJE, Ph.D. Texas A&M 1983. Inferences on interincome inequality parameters of a joint distribution of components of income and commodity expenditure.

BINDIGANAVALE N. SRINIDHI, Ph.D. Columbia 1984. Probability modeling of internal control systems in audit decision making.

STEPHEN P. STAGEBERG, Ph.D. Georgetown 1984. The impact of exogenous and endogenous technological change on labor productivity in the U.S. steel and textile industries between 1956 and 1978 using a translog production function.

JAMES H. STOCK, Ph.D. California-Berkeley 1983. Dynamic economic models subject to time deformation.

MOHAMMAD A. TAHER, Ph.D. Alberta 1983. An econometric study of energy and resource use in Canadian manufacturing industries.

CHU-WEI TSENG, Ph.D. Oklahoma State 1984. Some efficiency effects of public and private transfers: A cross-sectional analysis.

SHAWNEE K. VICKERY, Ph.D. South Carolina 1983. Multistage lot sizing for serial production.

FELIPE VILA GONZALEZ, Ph.D. Texas-Austin 1983. An optimal control model for policy analysis of the Mexican economy.

FREDERIK D. WIERSEMA, D.B.A. Harvard 1983. Price-cost dynamics: An empirical study.

LINDA P. WOODLAND, Ph.D. Purdue 1983. Development of a cross-sectionally constrained random coefficient model with an application to the estimation of the vintage effect on earnings profiles.

THOMAS W. ZUEHLKE, Ph.D. Florida 1983. Econometric modeling of the SRC price expectations data.

Monetary and Fiscal Theory, Policy and Institutions

CHARLES L. BALLARD, Ph.D. Stanford 1983. Evaluation of the consumption tax with dynamic general equilibrium models.

SHAROKH R. BASTANI, Ph.D. Rice 1984. Testing incidence of a differential tax on income from capital in the United States.

LEE A. BENDER, Ph.D. Temple 1984. Money stock control in Canada: Implications for prediction and policy.

LAXMI C. BHANDARI, Ph.D. Chicago 1984. Capital market imperfections, risk, and expected returns: Empirical evidence and some propositions.

EMANUAL S. BURSTEIN, Ph.D. Boston College 1984. An econometric model of state government capital spending.

SAMUEL R. CARTER, Ph.D. Tennessee 1984. Performance and adequacy of a state tax system: The case of Tennessee.

ABDUR R. CHOWDHURY, Ph.D. Kentucky 1983. The sectoral impact of monetary and fiscal policy.

GLEN H. COPPLESTONE, Ph.D. Western Ontario 1983. Implications of Canadian oil tax policies.

JOAQUIN COTTANI, Ph.D. Yale 1984. Disinflation policy in financially open economies.

CARLOS E. CUEVAS, Ph.D. Ohio State 1984. Intermediation costs and scale economies of banking under financial regulations in Honduras.

PHILIPPA S. DEE, Ph.D. Simon Fraser 1983. Optimal exchange market intervention.

GLADYS A. DEJESUS, Ph.D. Ohio State 1983. Monetary control in the Dominican Republic, 1965–80.

SATYA B. DUTTA, Ph.D. Clark 1984. Measuring fiscal influence: The full employment surplus vs. the cyclically neutral budget.

CLAUDIA H. DZIOBEK, Ph.D. Massachusetts 1984. International commercial bank lending to Third World countries, 1970–81: An analysis of U.S. bank behavior.

DONALD A. ELLIOTT, Ph.D. Georgia 1983. St. Louis Fed's monetary model, whence it came—how it thrived, 1970–82.

ELSSAYED M. ELSSAMADISY, Ph.D. North Carolina State 1983. The effects of credit constraints in dynamic optimization by consumers.

ZIBA FARHADIAN, Ph.D. George Washington 1984. The government deficit, the balance of payments, and role of financial development.

JOHN L. FAULKNER, Ph.D. Colorado-Boulder 1984. The effect of spending limitations on the allocation of public spending and an empirical study of Colorado's 7 percent spending limitation.

MICHAEL R. FEDEROW, Ph.D. Massachusetts 1983. Taxation and public school finance reform in Connecticut.

EDUARDO FERNANDEZ, Ph.D. Illinois-Urbana 1983. State intervention, public finances, and economics.

WILLIAM G. FOOTE, Ph.D. Fordham 1984. Money and the short-run determination of stock price.

SCOTT J. FREEMAN, Ph.D. Minnesota 1983. Transaction costs and the optimal quantity of money.

GERTRUD M. FREMLING, Ph.D. California-Los Angeles 1983. Monetary policy under a gold standard: How effective is it?

DOUG J. GILMORE, Ph.D. Illinois-Urbana 1983. The effects of taxation on corporate investment and financial policy.

DENNIS C. GLENNON, Ph.D. Missouri-Columbia 1983. Private saving and ultrarationality: An empirical investigation.

GARY B. GORTON, Ph.D. Rochester 1983. Banking panics and bank suspension of convertibility.

PHILIP J. GROSSMAN, Ph.D. Virginia 1984. Property tax assessment bias: A study of the property tax as a user charge.

LOTHAR HARZDORF, Ph.D. State University New York-Stony Brook 1983. Growth characteristics of large U.S. bank corporations.

JAMES G. HOEHN, Ph.D. Virginia 1983. The monetary instrument and lagged reserve accounting.

SELIM JAHAN, Ph.D. McGill 1983. Money, income, and the balance of payments of Bangladesh: A macroeconomic model.

JOSEPH P. JOYCE, Ph.D. Boston 1984. Analysis of the goals of Canadian monetary policy, 1970–80.

NICHOLAS V. KARAMOUZIS, Ph.D. Pennsylvania State 1984. The role of "openness" in the conduct of U.S. monetary policy.

JOSEPH D. KELLEY, Ph.D. California-Berkeley 1983. Macroeconomic factors affecting asset prices.

JUN YOUNG KIM, Ph.D. Minnesota 1984. Taxation, equilibrium dynamics, and welfare.

JANE E. KNODELL, Ph.D. Stanford 1984. Financial structure and financial stability: The American West, 1815–45.

GEORGE W. KUTNER, JR., Ph.D. Northwestern 1984. An empirical analysis of risk premiums in the negotiable certificate of deposit market from 1978 to 1981.

PHILIP J. LANE, Ph.D. Tufts 1983. An analysis of the municipal bond market.

ALFREDO M. LEONE, Ph.D. Minnesota 1983. Public finance and the behavior of inflation rate and relative commodity prices.

FREDERICK V. LINSE, Ph.D. Minnesota 1984. Implications of price risk for price support loan demand.

M. KEVIN MCGEE, Ph.D. Ohio State 1983. Taxation and dividend policies with stochastic earnings.

ALEX S. MACNEVIN, Ph.D. McMaster 1984. Fiscal integration and subcentral public sector inducements to Canadian interprovincial migration.

IQBAL MANSUR, Ph.D. Cincinnati 1983. Joint determination of asset and liability structure of the banking firm.

EVA MARIKOVA, Ph.D. Princeton 1983. The determination and pricing of the loan-to-value ratio in mortgage lending.

JOHN E. MARTINEZ, Ph.D. Oklahoma 1984. An empirical test of the impact of potential competition on banking structure.

PAUL M. MASON, Ph.D. Texas-Austin 1984. An alterna-

tive assessment of economies of scale in the Texas savings and loan industry.

ERNESTO G. MAY-KANOSKY, Ph.D. Massachusetts Institute of Technology 1984. A computational approach to optimal taxation and public pricing policies: The case of Mexico.

KAZUO MINO, Ph.D. Brown 1984. Growth, stability, and money supply policies.

MARK A. NADLER, Ph.D. Iowa State 1983. An analysis of some of the issues raised in the liquidity-preference loanable funds interest rate controversy.

LYNN A. NEJEZCHLEB, Ph.D. Nebraska-Lincoln 1982. Bank performance in meeting local community needs: A comparative study of independent chain banking in rural Nebraska.

CHU NGUYEN, Ph.D. Cincinnati 1983. A portfolio equilibrium approach to the theory of savings and loan associations.

JAMES A. OVERDAHL, Ph.D. Iowa State 1984. The CD futures market: Hedging and price discovery performance.

SANG Y. PARK, Ph.D. New York 1984. Taxes and equilibrium: Related essays.

JAMES W. PAULSEN, Ph.D. Iowa State 1983. Effects of license fees, unit, and ad valorem excise taxes on efficiency and maximum potential tax collections in perfectly competitive and monopolistically competitive industries.

GEORGE G. PENNACCHI, Ph.D. Massachusetts Institute of Technology. An analysis of alternative deposit insurance systems.

ANNETTE B. POULSEN, Ph.D. Ohio State 1983. The impact of Japanese and U.S. financial conditions on the activity of Japanese banks in the United States.

MITCHELL A. POST, Ph.D. Wisconsin-Madison 1984. An analysis of an economy with tax arbitrage: Consequences for saving and investment and vertical equity.

PETER C. RANGAZAS, Ph.D. Indiana-Bloomington 1983. Financial determinants of investment.

COSTAS RANOS, Ph.D. Indiana-Bloomington 1983. Conflict of state-local preferences: A control and a tax issue in a public good—the Indiana case.

SRIKALA RAVIKUMAR, Ph.D. Columbia 1984. Short-term financial capital: A model for the United States, 1961–71.

RITA L. REICHER, Ph.D. Yale 1983. The impact of federal grants-in-aid on local government and decision making.

BRADFORD G. REID, Ph.D. Toronto 1984. Government deficits and economic activity: Empirical evidence on the Ricardian equivalence hypothesis.

WILLIAM J. ROBINSON, Ph.D. Colorado-Boulder 1983. An analysis and alternative approach to the taxation of U.S. multinational corporations.

CHAIPAT SAHASAKUL, Ph.D. Rochester 1984. The measure and predictability of the average marginal tax rate.

JACQUES A. SCHOUPPE, Ph.D. Northwestern 1984. Subordinated capital augmentation debentures: A source of capital savings and loan associations.

ANDREW SENCHAK, Ph.D. Columbia 1983. The U.S. capital accumulation: 1963–77.

ROBERT W. STONE, Ph.D. Purdue 1983. The impact of institutions upon the pattern of public expenditures.

KONG YAM TAN, Ph.D. Stanford 1984. Flexible exchange rates and interdependence: Empirical implications for U.S. monetary policy.

POCHARA THEERATHORN, Ph.D. Northwestern 1984. The risk structure of interest rates.

ROBERT W. TURNER, Ph.D. Massachusetts Institute of Technology 1984. The effect of taxes on the form of compensation.

CHARLOTTE A. L. TWIGHT, Ph.D. Washington 1983. Government manipulation of constitutional-level transaction costs: An economic theory and its application to off-budget expenditure through the federal financing bank.

CHRISTIAN G. VAN WIJK, Ph.D. Columbia 1984. The role of trust and shared understandings in loan decisions.

ANDREW M. WELKI, Ph.D. Pennsylvania State 1983. Tax exporting and the degree of reliance on local tax bases.

STEPHEN D. WILLIAMSON, Ph.D. Wisconsin-Madison 1984. Four essays on financial intermediation.

SUK HEUN YOON, Ph.D. Northwestern 1984. Essays on bank lending.

CHORNG-JIE AUGUSTINE YU, Ph.D. Washington-St. Louis 1984. Evaluation of Federal Reserve policies on short-run monetary control.

International Economics

RANDA AZAR, Ph.D. Southern California 1983. Ex. .ining the portfolio balance theory of foreign exchange rates: A time-series analysis.

HERBERT L. BAER, JR., Ph.D. Northwe rn 1984. Competition among the multinationals. An examination of equilibrium in multimarket oligopolies.

TAEHO BARK, Ph.D. Wisconsin-Madison 1983. A dynamic analysis of macroeconomic policy changes in a small open economy facing an international capital market.

ANTHONY G. BARRETT, Ph.D. George Washington 1984. The determinants of international debt rescheduling by time, region, and stage of development: An empirical analysis.

FRANCISCO M. BATALLER, Ph.D. Ohio State 1983. The determinants of import protection in a developing country: Spain in the 1960's.

DAVID A. BRODSKY, Ph.D. Massachusetts Institute of Technology 1983. Essays in international economics.

JACQUELINE M. BRUX, Ph.D. Michigan 1983. Economic

analysis of buffer stocks to stabilize the international grain markets.

TAIN JY CHEN, Ph.D. Pennsylvania State 1983. Foreign investment, duty-free zones, and national welfare.

ISAIAS COELHO, Ph.D. Rochester 1984. Welfare aspects of factor movements and trade.

SUSAN M. COLLINS, Ph.D. Massachusetts Institute of Technology 1984. Fixed exchange rates, devaluations, and credibility crises.

DEBORAH J. CUNNINGHAM, Ph.D. Kentucky 1983. On forecasting exchange rates.

ANDRES D. DAUHAJRE, Ph.D. Columbia 1983. Dual exchange rates dynamics with incomplete segmentation and rational expectations.

CASPER G. DE VRIES, Ph.D. Purdue 1983. Three essays on the interaction of international asset and commodity trade.

HUBERTA DEWITTY, Ph.D. Stanford 1984. The role of commercial bank positions in foreign exchange markets.

STEVEN V. DUNAWAY, Ph.D. George Washington 1984. A model of U.S. portfolio capital flows.

GREGORY DYBALSKI, Ph.D. George Washington 1984. An economic critique of the international tin agreement's market control mechanisms.

CHARLES M. ENGEL, Ph.D. California-Berkeley 1983. Essays in the asset-market approach to exchange rates.

ALBERTO GIOVANNINI, Ph.D. Massachusetts Institute of Technology 1984. Essays on exchange rates and asset markets.

LAWRENCE D. GWINN, Ph.D. Kansas 1983. The transmission of U.S. income disturbances to Canada under flexible and fixed exchange rates.

DANIEL HIMARIOS, Ph.D. Virginia Polytechnic Institute 1984. Devaluation, the trade balance, and the capital account: Theory and evidence.

KENJIRO HIRAYAMA, Ph.D. Yale 1983. Effectiveness of stabilization policy under flexible exchange rates.

CHRISTOPHER B. HODGES, Ph.D. Minnesota 1983. Low-income countries and the world wheat trade.

FRANK T. HOLLENBECK, Ph.D. South Carolina 1984. A portfolio equilibrium approach to exchange rate determination.

HONG HWANG, Ph.D. Queen's 1983. Three essays on international trade with imperfect competition.

HOJIN KANG, Ph.D. Ohio State 1983. Import competition and its impact on domestic industries.

JUNG MO KANG, Ph.D. Northern Illinois 1983. The profit function approach to the estimation of import demand and export supply: The cases of Korea and Japan.

ARLEN R. KANGAS, Ph.D. Washington State 1984. The effect of log export restrictions on employment: A test of the Stolper-Samuelson theorem.

KYUNG SOO KIM, Ph.D. Pennsylvania 1984. Two essays on international monetary economics in an overlapping generations model.

JOHN H. KITCHEN, Ph.D. Pittsburgh 1983. Money supply announcements and the foreign exchange market: Evidence on the Fisher effect and market efficiency.

AKIRA KOHSAKA, Ph.D. Stanford 1984. Determination of the yen-dollar exchange rate, 1973–81.

ROBERT KORAJCZYK, Ph.D. Chicago 1983. The pricing of forward and futures contracts for foreign exchange.

ERCAN M. KUMCU, Ph.D. Boston 1983. Exchange rate determination and welfare implications of commercial and monetary policies.

RAJIV LALL, Ph.D. Columbia 1984. Multinationals from LDCs: The competitive advantage of Indian firms investing abroad.

WON-YOUNG LEE, Ph.D. Michigan 1983. A theory of royalty payment and international technology transactions.

JOHN LEONARD, Ph.D. State University of New York-Stony Brook 1983. An empirical study of speculation in the yen/dollar exchange rate, 1974–77.

WEI L. LOH, Ph.D. Tufts 1984. Foreign technical assistance to Malaysia: Some World Bank road and road transport projects.

MICHEL G. MAILA, Ph.D. Chicago 1983. Systematic exchange risk and the volume of international trade under pegged and floating exchange rates.

CATHERINE L. MANN, Ph.D. Massachusetts Institute of Technology 1984. Trade and finance relations between the developed and the newly industrialized countries.

MOHAMMAD E. MASSOUDNIA, Ph.D. Nebraska-Lincoln 1983. An empirical study of monetary approach to the balance of payments: The case study of Iran during 1962–78.

JAIME MARQUEZ, Ph.D. Pennsylvania 1983. International transmission of oil price effects and the optimal derivation of oil prices.

STEVEN J. MATUSZ, Ph.D. Pennsylvania 1983. Technological uncertainty, unemployment, and the pure theory of international trade.

PANOS D. MAVROKEFALOS,, Ph.D. California-Los Angeles 1984. Determinants of foreign direct investment: The role of post-World War II technological process: Case study of the semiconductor industry.

ROBERT N. MEFFORD, Ph.D. California-Berkeley 1983. Determinants of labor productivity differences in industrial plants: An international study.

E. HUBERT MEHNERT, Ph.D. Utah 1984. Controlling Eurodollar flows: The German experience, 1960–68.

MASSOUD MOKHTARI SAGHAFI, Ph.D. Southern California 1983. Essays on the theory of international economics and welfare resource transfers, trade priorities, and regional welfare.

ROBERT G. MURPHY, Ph.D. Massachusetts Institute of Technology 1984. Essays on macroeconomic adjustment in open economies.

AN VINH NGUYEN, Ph.D. Pittsburgh 1984. The U.S. government intervention in the foreign exchange market.

FARHANG NIROOMAND, Ph.D. Michigan State 1983. Determinants of the structure of U.S. foreign trade in manufacturing, 1963–80.

PANTELIS PANTELIDES, Ph.D. Wayne State 1983. Trade liberalization between the United States and less developed countries: Testing the U.S. generalized system of preferences.

RODRIGO PAROT, Ph.D. Boston 1984. Export responses to trade liberalization: A case study of some Chilean exports.

BRIAN R. PINTO, Ph.D. Pennsylvania 1983. On oligopolistic dumping, and essays in time-series.

GERALD V. POST, Ph.D. Iowa State 1983. Optimal tariffs and retaliation with perfect foresight.

GARY M. QUINLIVAN, Ph.D. State University of New York-Albany 1983. A cross-spectral analysis of exchange rate volatility during the 1970's and 1980's.

MOHAMMAD T. RAMTOOLAH, Ph.D. McGill 1983. The behavior of foreign exchange markets: An empirical study.

ANNE RICHARDS-LOUP, Ph.D. George Washington 1984. Multilateral trade in manufactured goods: A South-South, South-North comparison.

MIGUEL A. RODRIGUEZ, Ph.D. Yale 1983. Inflation, the balance of payments, and real output in Venezuela.

FRANCIS J. ROZWADOWSKI, Ph.D. Columbia 1984. Alternative monetary strategies in an open economy.

ZAIDI SATTAR, Ph.D. Boston 1984. Resource allocation impacts and activation effects of trade policies.

ANDREAS SAVVIDES, Ph.D. Florida 1983. Bank loan rate indexation in the Eurodollar markets.

ERIK N. SCARSBOROUGH, Ph.D. Hawaii-Manoa 1983. Relative effects of alternative export and industrial promotion schemes in the Philippines and Thailand.

DAVID C. SCHIRM, Ph.D. Pennsylvania State 1983. International investor behavior and the implementation of Federal Reserve domestic monetary policy: The floating exchange rate period.

HAMID SHAHRESTANI, Ph.D. Cincinnati 1984. The impact of U.S. foreign direct investment on Canadian economy.

CATHERINE-MARGUERITE T. SHALEN, Ph.D. Columbia 1984. Hedging in the silver futures and forward markets: Theory and practice.

PETER A. SHARP, Ph.D. Claremont 1984. Forward exchange risk premia in efficient markets.

THOMAS E. SHEETZ, Ph.D. Texas-Austin 1983. The International Monetary Fund, its monetarist model, and Peruvian debt crises.

ANDRES SOLIMANO RATINOFF, Ph.D. Massachusetts Institute of Technology 1984. Devaluation, unemployment, and inflation: Essays on macroeconomic adjustment.

OLIVER CHIEN-CHIH SOONG, Ph.D. Colorado State 1983. Impact of devaluation on trade balance: A theoretical analysis and the empirical evidence of Taiwan.

BARBARA J. SUNDQUIST, Ph.D. Colorado-Boulder 1984. Monetary and political influences on balances of payments and exchange rates and their redistributive effects.

BHASU B. SUPAPOL, Ph.D. McGill 1983. International liquidity, reserves, and monetary gold.

PRASAS TANGMATITAM, Ph.D. Hawaii 1983. The macroeconomic adjustment of the balance of payments of Asian countries.

EVAN S. THOMAS, Ph.D. California-Davis 1984. Sovereign debt crises.

ROSEMARY THOMAS, Ph.D. Fordham 1984. The monetary approach to exchange market pressure: An application to the Peruvian experience.

BANGORN TUBTIMTONG, Ph.D. Colorado-Boulder 1983. Time-series response pattern and the test of foreign exchange market efficiency.

JESUS G. TURRUBIATE, Ph.D. Notre Dame 1984. Trade imbalance, income distribution, and employment problems in the Mexican economy.

PORNPEN VORASOPONTAVIPORN, Ph.D. Notre Dame 1984. Trade policy for employment and income distribution: The case of Thailand.

JAMES D. WHITNEY, Ph.D. Wisconsin-Madison 1983. International oligopoly—Cournot behavior and commodity trade: A theoretical and empirical investigation.

BETH V. YARBROUGH, Ph.D. Washington 1983. The international transmission of economic disturbances: Information and the choice of exchange rate regime.

GIRMA ZELLEKE, Ph.D. Colorado-Boulder 1984. Deviation of exchange rates from purchasing power parity.

ZINSER, LEE D., Ph.D. Iowa State 1983. Three essays in international and intraregional economics.

Business Administration; including Business Finance and Investment, Insurance Marketing, and Accounting

MADELYN ANTONCIC, Ph.D. New York 1983. Shifts in equilibrium stock prices due to information shocks and high and volatile real interest rates: Where does the Fed fit in? Two problems in the measurement of imperfectly observed time varying parameters.

DURWOOD R. BAGBY, Ph.D. South Carolina 1983. A strategic management approach to policy analysis.

ARVIND BHAMBRI, D.B.A. Harvard 1984. The internal dynamics of corporate responsiveness to public policy issues: An exploratory study.

JOHN J. BINDER, Ph.D. Chicago 1983. Measuring the effect of regulation with stock price data: A new methodology.

ALLYN G. BLENCOE, Ph.D. South Carolina 1984. An

evaluation of raters' cognitive processes in performance appraisal: A laboratory study.

VICTOR M. BORUN, Ph.D. New York 1984. The cost of issuance of new equity: A case of the utility industry.

THEODORE BYRLEY, Ph.D. Cincinnati 1983. Changing patterns of the security market line, 1952 to 1984.

STEVEN A. CARVELL, Ph.D. State University of New York-Binghamton 1984. The impact of analyst neglect on stock market performance.

SUSAN J. CHAPLINSKY, Ph.D. Chicago 1984. The economic determinants of leverage: Theories and evidence.

SRIS CHATTERJEE, Ph.D. Columbia 1983. An investigation into corporate debt policy.

DAVID C. CHO, Ph.D. New York 1983. Essays in finance and operations research.

KENNETH S. CHOIE, Ph.D. Michigan 1984. The optimal offering price of new issues of corporate securities.

THOMAS N. CLOUGH, D.B.A. Harvard 1983. New product development in a technology-based business: An exploratory study.

H. ALLAN CONWAY, D.B.A. Harvard 1983. The evolution of the role of the corporate office in the diversified firm.

MARGARET R. CVAR, D.B.A. Harvard 1984. Competitive strategies in global industries.

THEODORE E. DEGNAN, Ph.D. South Carolina 1983. A computer-coded content analysis study of tax cases involving valuations of large holdings of public traded stock.

PETER D. EASTON, Ph.D. California-Berkeley 1984. Empirical aspects of an information perspective on accounting.

PETER J. ELMER, Ph.D. Purdue 1983. Relative prices, real activity, and the valuation of shares.

MUSTAFA N. GULTEKIN, Ph.D. New York 1984. Essays in finance.

EITAN GUREL, Ph.D. California-Berkeley 1984. On the interest rate risk of financial institutions.

MAHMOUSD HADDAD, Ph.D. Alabama 1984. Comparative empirical study on beta stationarity based on the mean-semivariance and the mean-variance capital asset pricing models.

PHILIPPE C. HASPESLAGH, D.B.A. Harvard 1983. Portfolio planning approaches and the strategic management process in diversified industrial companies.

CAROLINE M. HENDERSON, D.B.A. Harvard 1983. The effect of price promotion on coupon redemption and brand choice.

JULIE H. HERTENSTEIN, D.B.A. Harvard 1984. Innovation in management control: A comparative study of inflation-adjusted accounting systems in diversified firms.

SHARON M. HOCKETT, D.B.A. Harvard 1983. Prescriptive interactions in decision making under uncertainty: The case of clinical medicine.

SUNGBIN C. HONG, Ph.D. California-Berkeley 1984. The outcome of bankruptcy: Model and empirical test.

KEITH B. JARRETT, D.B.A. Harvard 1983. Market expectations of riskless interest rates: Measurements and uses.

MICHAEL D. JOHNSON, Ph.D. Chicago 1983. Decision processing and product comparability: A theory of strategy selection.

BRUCE L. JONES, Ph.D. Illinois 1984. Capitalization methods of production credit associations and their impact on borrowers' costs.

PHILIPPE JORION, Ph.D. Chicago 1983. Portfolio analysis of international equity investments.

MICHAEL A. KAMINS, Ph.D. New York 1984. The impact of involvement, advertising type, and expectation level on product evaluation.

DONALD KEIM, Ph.D. Chicago 1983. The interrelation between dividend yields, equity values, and stock returns: Implications of abnormal January returns.

ALAN W. KLEIDON, Ph.D. Chicago 1983. Stock prices as rational forecasters of future cash flows.

APRIL M. KLEIN, Ph.D. Chicago 1983. Voluntary corporate divestitures: Motives and consequences.

MARK S. KLOCK, Ph.D. Boston College 1983. Dividend and debt decisions of firms.

ROBERT J. KOPP, D.B.A. Harvard 1984. Improving sales force effectiveness in trade promotions: An exploratory study.

DAVID J. LEATHAM, Ph.D. Purdue 1983. Optimal choice of fixed and variable interest rate loans.

DAL-HEE LEE, Ph.D. Pennsylvania 1983. Three essays on the efficiency and volatility of asset prices.

DAVID MAZURSKY, Ph.D. New York 1983. A behavioral process approach for tracking store image development.

RICHARD MOLZ, Ph.D. Massachusetts 1984. Corporate boards of directors: A study of the relationship of board structure and composition and corporate performance.

DEB K. NAG CHOWDHURY, Ph.D. Simon Fraser 1983. An exploratory study of manager-developed decision support systems.

MARILYN NEIMARK, Ph.D. New York 1983. The social construction of annual reports: A radical approach to corporate control.

DONALD G. NORRIS, Ph.D. California-Berkeley 1983. An extension and validation of the cognitive response model.

TOM W. OGLESBEE, JR., Ph.D. South Carolina 1983. Managerial evaluations of accounting information system success: An empirical study.

KIRK L. PHILIPICH, D.B.A. Indiana 1984. The effect of three factors on the accuracy of univariate time-series forecasts of quarterly earnings.

MORAGEA Y. SALEM, Ph.D. South Carolina 1984. An empirical analysis of operational planning in small

businesses.

ALEXANDER J. SANNELLA, Ph.D. New York 1984. Segment reporting: The cost allocation issue.

ALICE M. SAPIENZA, D.B.A. Harvard 1984. Believing is seeing.

FRIEDA W. SHAVIRO, Ph.D. Fordham 1984. An analysis of cash and futures prices in the delivery period of maturing contracts in the coffee "C" market, 1972–81.

SOUMYENDRANATH SIKDAR, Ph.D. Minnesota 1984. The economics of labor-managed enterprises.

TOM STOBER, Ph.D. Chicago 1983. Nominal contracting and accounting policies: Effects on the valuation of rate-regulated firms.

SENYO YAWO TSE, Ph.D. California-Berkeley 1983. Security valuation, accounting information, and the usefulness of segment data.

DOUNG-VAN VU, Ph.D. Chicago 1984. An examination of corporate nonconvertible bond calls and their impact on security prices.

GREGORY B. WAYMIRE, Ph.D. Chicago 1984. An empirical analysis of voluntary management earnings forecasts.

EPHRAIM WEINGARTEN, Ph.D. New York 1983. Manner of use and mode of presenting celebrities in advertising: An attribution theory approach.

PAULA E. WIEHRS, Ph.D. South Carolina 1983. An empirical study of capital investment incentives in an inflationary environment.

THOMAS A. WRIGHT, Ph.D. California-Berkeley 1984. Examination of the processes of cognitive appraisal coping and adaptation in an organizational field setting.

AL-RAI K. ZIAD, Ph.D. New York 1983. The impact of corporate forecast's accuracy of earnings per share on the stock market: Empirical analysis.

Industrial Organization and Public Policy: including Economics of Technological Change, and Industry Studies

DAVID AJAKAIYE, Ph.D. Boston 1984. Public enterprise price liberalization policy under alternative fiscal policy regimes: An input-output model of the short run impacts with simulation for Nigeria.

DONALD ALEXANDER, Ph.D. Pennsylvania State 1983. The competitive relationships among multibank holding companies across geographic markets: A test of the mutual forebearance hypothesis.

MOHAMMAD S. AL SABBAN, Ph.D. Colorado-Boulder 1983. Viability of industrial integration within the Gulf Corporation Council: The case of petrochemical industries.

JAMES L. ANDERSON, Ph.D. California-Davis 1983. Bioeconomic interaction between aquaculture and the common property fishery with application to Northwest salmon resources.

ANDREAS ANTONIOU, Ph.D. McGill 1983. Ownership and control of the largest Canadian-owned corporations, 1979.

SANDRA O. ARCHIBALD, Ph.D. California-Davis 1984. A dynamic analysis of production externalities: Pesticide resistance in California cotton.

JAMES R. ASHLEY, Ph.D. Wisconsin-Madison 1984. The effects of regulation on the diffusion of technological innovations in the U.S. hospital industry.

AVIVA A. BERK, Ph.D. Columbia 1984. Cost-effectiveness analysis of ambulatory vs. inpatient care.

JAMES A. BERKOVEC, Ph.D. Massachusetts Institute of Technology 1984. Automobile market equilibrium.

GERMAN BOTERO-ARBOLEDA, Ph.D. Boston 1984. Transport firm behavior under demand uncertainty: The case of Colombia.

EDWARD C. BOYER, Ph.D. Temple 1984. The role of sunk costs in a theory of dynamic industry structure.

JULIE CASWELL, Ph.D. Wisconsin-Madison 1984. Direct stockholding and network interlocking directorate control of firms in agribusiness sector, 1976.

ANDREW J. CHALK, Ph.D. Washington-St. Louis 1983. The economics of airline safety.

WILLIAM F. CHAPPELL, Ph.D. South Carolina 1983. A simultaneous equations analysis of market structure, conduct, and performance.

CHEE WAH CHEAH, Ph.D. Hawaii-Manoa 1984. Economic perspective on the regulation of transborder data flow.

KWO-SHIN CHEN, Ph.D. Purdue 1984. The growth of large cooperative and proprietary firms in the U.S. food sector.

SHIN-HWAN CHIANG, Ph.D. State University of New York-Buffalo 1983. Theoretical issues in managerial compensation.

DAVID KENSU CHIN, Ph.D. George Washington 1983. The effect of a subsidy on the receiver's resource constraint: The case of transit capital grants.

SANG MOK CHOI, Ph.D. Hawaii-Manoa 1984. Cross-sectional analysis of demand for laborer and capital inputs in the manufacturing industries: A case study of Korea.

CHINSEUNG CHUNG, Ph.D. Georgia 1983. Performance evaluation of public enterprise.

LAWRENCE T. CLARK, JR., Ph.D. Kansas State 1984. The effects of government policy and external shocks on U.S. regional production and trading patterns of petroleum and petroleum products, 1974–82.

BEN S. COLLIER, Ph.D. Purdue 1984. A theoretical and empirical investigation of contractual provisions which inhibit franchisee competition after termination.

CATHLEEN J. COOLIDGE, Ph.D. Virginia Polytechnic Institute 1984. The economics of firm size, market structure, and the market for political influence.

LAURENCE D. CORNELL, Ph.D. Michigan State 1983. Implications of structural change in U.S. demand

for meats on U.S. livestock and grain markets.

PAULO C. COUTINHO, Ph.D. Pennsylvania 1984. On the microeconomics of contracts.

SUELLEN M. CURKENDALL, Ph.D. Virginia 1984. Theory of the firm facing price controls: A rational expectations approach.

ALAN J. DASKIN, Ph.D. Massachusetts Institute of Technology 1984. Essays on firm diversification and market concentration.

RAYMOND J. DENECKERE, Ph.D. Wisconsin-Madison 1983. Three essays in industrial organization.

JAMES E. FIELDS, Ph.D. Stanford 1983. Essays on the provision of quality, and information concerning quality in oligopolized and monopolized industries.

EDWARD C. GALLICK, Ph.D. California-Los Angeles 1984. Exclusive dealing, specialized assets, and joint ownership: A study of the tuna industry.

WILLIAM R. GATES, Ph.D. Yale 1984. Federal participation in commercial technology development: The case of solar commercial technologies.

LORI GERRING, Ph.D. Cincinnati 1984. Structural changes between models of fossil fuel demand by steam-electric power plants.

JEFFREY GOLDBERG, Ph.D. Pennsylvania 1983. A transactions cost analysis of franchising with empirical applications to the restaurant industry.

JOHN J. GONZALES, Ph.D. Wisconsin-Madison 1983. The effects of federal regulation on U.S. petroleum markets.

BURL W. HAAR, Ph.D. Nebraska-Lincoln 1984. Collusion in the highway construction industry: Economic evidence as an investigative tool.

BRIAN G. HANESSIAN, Ph.D. Princeton 1984. Three essays in industrial organization.

KEITH W. HEIMFORTH, Ph.D. Iowa State 1983. The effects of conglomerate mergers on concentration in manufacturing industries, 1967–77.

ROBERT M. HOUSE, Ph.D. Rice 1984. A mathematical programming model for agriculture sector policy analysis.

JOHN H. HUSTON, Ph.D. Wisconsin-Madison 1983. The cost structure of the regulated motor carrier industry.

FREDERICK B. JENNINGS, Ph.D. Stanford 1983. Public policy, organizational structure and response to threats of technological obsolescence: A postmortem on British canals.

ELIZABETH J. JENSEN, Ph.D. Massachusetts Institute of Technology 1984. Research expenditures, regulation, and innovation in the ethical pharmaceutical industry.

KI WON KANG, Ph.D. Rutgers 1983. Structural changes in production technology of the U.S. consumer electronics industry: 1958–80.

JOHN V. KIRNAN, Ph.D. Fordham 1984. A comparative analysis of the price-quality relation of American and foreign automobiles, 1971–80.

MARK N. KOSMO, Ph.D. Wisconsin-Madison 1983. Access to OCS drilling rights: Evaluating its determinants and their effects on the structure of the petroleum industry.

S. GANESH KRISHNAN, Ph.D. Nebraska-Lincoln 1984. Residential and industrial demand for electricity in Nebraska: An econometric study.

YOUN H. LEE, Ph.D. Wisconsin-Madison 1984. The effect of the new jobs tax credit on employment, price, and output of the manufacturing industries.

LODE LI, Ph.D. Northwestern 1984. A stochastic theory of the firm.

JOHNNIE B. LINN III, Ph.D. Texas A&M 1983. Property rights formation and enforcement in the competitive equilibrium as a function of input elasticities in the production and enforcement sectors.

RICHARD G. LINOWES, D.B.A. Harvard 1984. Strategy, structure and information processing: A study of apparel retailing.

JEANNE M. LOGSDON, Ph.D. California-Berkeley 1983. Organizational responses to environmental issues: Oil refining companies and air pollution.

ROBERT LOUBE, Ph.D. Michigan State 1983. The U.S. district heating industry: A case study of corporate strategy and public utility regulation.

VICKI A. MCCRACKEN, Ph.D. Purdue 1984. An econometric analysis of the away-from-home food market.

NICCIE L. MCKAY, Ph.D. Massachusetts Institute of Technology 1984. The economics of the medical diagnostic imaging equipment industry.

JAMES MACDONALD, Ph.D. State University of New York-Buffalo 1983. Effects of diversification in U.S. manufacturing: An empirical investigation.

FRANCO MALERBA, Ph.D. Yale 1983. The effects of vertical integration and public procurement policy on the evolution of the American, German, and Italian semiconductor industry.

WILLIAM N. MASS, Ph.D. Boston College 1984. The diffusion of technology: The case of the Northrop loom.

ELLEN M. MILLER, Ph.D. Florida 1983. The impact of the automatic fuel adjustment clause on production efficiency for electric utilities.

THOMAS M. MITCHELL, Ph.D. Brown 1984. An alternative model for accounting for technical progress: Theory and estimation.

FELIPE P. MONTT, Ph.D. Minnesota 1983. The world copper market.

SVERRE MUNCK, Ph.D. Stanford 1983. An energy policy simulation model of the Norweigan economy.

NANDU J. NAGARAJAN, Ph.D. Northwestern 1984. Auditor contracts and regulatory standards: An approach to investigating the effectiveness of legal liability in regulation of auditor performance.

CRAIG M. NEWMARK, Ph.D. California-Los Angeles 1984. A study of the incidence of horizontal price fixing in U.S. manufacturing industries.

MONICA G. NOETHER, Ph.D. Chicago 1983. The growing supply of physicians: Are the entry barriers breaking?

MICHAEL J. O'HARA, Ph.D. Nebraska-Lincoln 1983. The Nebraska Power Review Board: Regulating a publicly owned electric utility industry.

STEFAN REICHELSTEIN, Ph.D. Northwestern 1984. Information and incentives in economic organizations.

DAVID L. REIFSCHNEIDER, Ph.D. Wisconsin-Madison 1983. The forecasting of the distribution of weekly peak electricity demand and its application to planned outage scheduling.

JOAN E. RICART I COSTA, Ph.D. Northwestern 1984. On managerial contracts.

CRAIG R. ROACH, Ph.D. Wisconsin-Madison 1983. Coal use by industry and the associated air pollution emissions in the period from 1980 to 2000 under alternative market and regulatory conditions.

ANTHONY P. RODRIGUES, Ph.D. California-Berkeley 1984. Short-run potential output growth with putty and clay.

THOMAS W. ROEHL, Ph.D. Washington 1983. An economic analysis of industrial groupings in postwar Japan.

SUPAVUD SAICHEUA, Ph.D. Hawaii-Manoa 1984. Input substitution in Thailand's manufacturing sector and assessing implications for energy policy.

DEBORAH S. SCHELLENBERG, D.B.A. Indiana 1983. The effect of congruence among strategy, structure, and reward criteria on organizational performance.

PAULINE J. SHELDON, Ph.D. Hawaii-Manoa 1984. An analysis of tour packaging.

MARY C. SIMPSON, Ph.D. Claremont 1984. Determining optimal size in a final process: The case of electricity distribution.

RICHARD S. STEINBERG, Ph.D. Pennsylvania 1984. Two essays on the nonprofit sector.

PATRICIA L. SUCCAR, Ph.D. Northwestern 1984. Technical choice and the learning curve: A restatement of the infant industry argument.

VICTOR J. TREMBLAY, Ph.D. Washington State 1983. The efforts of firm behavior and technology on firm size: A case study of the U.S. brewing industry.

BIJAN VASIGH, Ph.D. State University of New York-Binghamton 1984. Demand for consumer durables under operating cost shocks: A study in the automobile market.

MICHAEL G. VITA, Ph.D. Wisconsin-Madison 1984. Profits, concentration, and the collusion-efficiency debate: A simultaneous equations model.

JOHN C. WELDING, Ph.D. Nebraska-Lincoln 1984. The impact of navigation user charges on the transportation of grain from Nebraska.

ROBERT L. WILLS, Ph.D. Wisconsin-Madison 1983. The impact of structure and advertising on brand pricing in processed food products.

ROBERT J. WINDLE, Ph.D. Wisconsin-Madison 1984. The cost structure of the U.S. urban bus transit industry.

Agriculture and Natural Resources

CICELY M. AMARAL, Ph.D. North Carolina State 1984. Information in U.S. Agriculture: The farm press.

NICOLE S. BALLENGER, Ph.D. California-Davis 1984. Agricultural policy analysis for Mexico: Sectoral and macro impacts.

MICHAEL L. BOYD, Ph.D. Stanford 1984. The private and social sectors in postwar Yugoslav agriculture: History, policy, and performance.

MING-HONG CHOW, Ph.D. Kansas State 1984. Economic impacts of structural changes in the wheat logistics system: Exporting winter wheat from twelve counties in northwest Kansas.

MOO NAM CHUNG, Ph.D. Missouri 1983. Review of Korean small farm economy with special emphasis on farm size questions, including economies of size in paddy production.

ROGER K. CONWAY, Ph.D. George Washington 1984. U.S. agriculture in the world economy: A stochastic coefficients approach.

PAMELA M. J. COX, Ph.D. Tufts 1984. Energy efficiency in agricultural production: A case study of Senegal.

AUGUSTO C. S. DE OLIVEIRA, Ph.D. Vanderbilt 1983. An application of the theory of the farm-household: Small farms and sharecroppers in northeast Brazil.

ROBERT DEUSON, Ph.D. Minnesota 1983. The lower Moulouya irrigation project: An assessment of its socioeconomic impact on the northeast of Morocco.

ARNE DISCH, Ph.D. Yale 1983. Agricultural policies and real income changes: An application of duality theory to Brazilian agriculture.

DANIEL L. DUNN, Ph.D. New Hampshire 1983. An econometric analysis of U.S. farmland prices, 1940–80.

JAMES S. EALES II, Ph.D. California-Davis 1983. Modeling searching behavior in the pink shrimp fishery: Area choice and information gathering.

MARTIN L. FISCHER, Ph.D. Minnesota 1984. Financing agricultural cooperatives: Economic issues and alternatives.

JOHN R. FISKE, Ph.D. Iowa State 1983. An analysis of the impact of the down payment ratio and farm financial structure on the selling price of Iowa farmland.

GWENDOLYN G. FLOWERS, Ph.D. Iowa State 1983. An economic analysis of the relation of farm land values and returns.

DAFIE'A A. GABIR, Ph.D. Ohio State 1983. Assessing organizational effectiveness: Two Sudanese agribusiness cases.

THOMAS E. GILLARD-BYERS, Ph.D. Illinois 1984. The

allocation of resources in the nonmechanized small farm household of Dumila, Mkundi, and Magole villages and the Kilosa district, Tanzania: A nutrition-based approach.

LUIZ R. GRACA, Ph.D. Purdue 1983. The effect of inflation on capital formation in agriculture: A study of tractor sales in the United States.

LEONARDO E. GREEN, Ph.D. Missouri 1983. A residual supplier model of international trade in corn.

DAVID HENNEBERRY, Ph.D. Wisconsin-Madison 1984. The capitalization of exclusive agricultural zoning into Wisconsin farmland values.

ROGER S. HEWETT, Ph.D. Illinois-Urbana 1984. Natural resource economic theory: An implication to Illinois coal.

ROBERT A. HILL, Ph.D. Minnesota 1984. An economic analysis of rice production in the Dominican Republic.

HSIN-HUI HSU, Ph.D. Michigan State 1984. A translog cost function study of purchased inputs used in U.S. farm production.

RAY G. HUFFAKER, Ph.D. California-Davis 1983. The distribution of economic rents when irrigated farmland is leased.

ALICE M. HUGHEY, Ph.D. Columbia 1984. A model of joint resource extraction: Natural gas and helium.

DONALD J. JANKOWSKI, Ph.D. Purdue 1983. On search for mineral resources: Planning and implementing an exploration program.

NIK F. KAMIL, Ph.D. Washington State 1984. An economic analysis of the Malaysia rice sector: Prospects and policy alternatives.

AAASHA KAPUR, Ph.D. Iowa State 1983. Application of sustainable agricultural production policies by means of a recursive econometric programming model.

BARBARA KATES-GARNICK, Ph.D. Tufts 1984. Consumer country cooperation in energy: Some limits and possibilities.

DANIEL E. KAUFFMAN, Ph.D. Michigan State 1983. A pork contract market: An investigation into attitudes about and possibilities of such a market for slaughter hogs.

ABBAS A. KAZEMI, Ph.D. Illinois 1983. Economic analysis of modifying cropping patterns in selected regions in Iran.

SOHEILA KHOII HASSANZADEH BORESHI, Ph.D. California-Davis 1984. The economic impact of alternative acreage limitations on the income and employment of farms and rural communities: The case of Westlands Water District, California.

ANDREW L. LARKIN, Ph.D. Nebraski-Lincoln 1982. Extension of the theory of institutionalism to the problem of environmental disruption.

PIERRE H. LEMOINE, Ph.D. Stanford 1984. Water resources management in the Salinas Valley: Integration of economics and hydrology in a closed control model.

HSIANG-HSI LIU, Ph.D. Iowa State 1983. An annual simultaneous equations econometric model of U.S. corn and soybean cash and futures markets.

JOHN R. LIVERNOIS, Ph.D. British Columbia 1984. An economic model of oil extraction: Theory and estimation.

MARIE L. LIVINGSTON, Ph.D. Colorado State 1984. A conceptual model of institutional innovation and its application to natural resource policy.

EDWARD W. MCLAUGHLIN, Ph.D. Michigan State 1983. Buying and selling practices in the fresh fruit and vegetable industry: Implications for vertical coordination.

PATRICIA E. MCLEAN, Ph.D. Ohio State 1984. A quadratic interregional and intertemporal model for grain marketing in Ohio.

NIK H. MUSTAPHA, Ph.D. Washington State 1984. Economic analysis of management alternatives for Malaysian fisheries.

PRANTASH NAG, Ph.D. Minnesota 1983. A study of apparent irrationality in the production behavior of rice farmers in West Bengal, India: A case study.

PLINIO M. NASTARI, Ph.D. Iowa State 1983. The role of sugar cane in Brazil's history and economy.

DENNIS L. NEF, Ph.D. Iowa State 1983. Efficiency of soil conservation investments and economic implications for change.

MICHAEL L. NIEWSIADOMY, Ph.D. Texas A&M 1983. Adjusting to diminishing water supplies in irrigated agriculture.

HEE MOCK NOH, Ph.D. Kansas State 1983. Prospects of grain production, consumption, and trade in China.

MARK O. ODHIAMBO, Ph.D. California-Davis 1983. Production risk and decision making: Testing alternative econometric models with evidence from Egyptian cotton production.

PHILIP L. PAARLBERG, Ph.D. Purdue 1984. Endogenous policy formation in the imperfect world wheat market.

ISABELITA M. PABUAYON, Ph.D. Kansas State 1983. Econometric analysis of the Philippine sugar economy and evaluation of sugar policies.

BONNIE SALIBA, Ph.D. Wisconsin-Madison 1983. An economic analysis of the relationship between soil conservation behavior and farmland characteristics.

KURT C. SCHAEFER, Ph.D. Michigan 1984. A portfolio model of mixed farming in less developed countries, with an application to project evaluation in southeast Niger.

ROBERT J. SHEVLIN, Ph.D. Columbia 1984. Economic consequences of the variability effect in solar thermal irrigation.

ANDREW B. SISSON, Ph.D. Tufts 1983. Economic im-

plications of rural development projects: improved cotton and irrigated rice in North Cameroon.

SOEMIRAT SLAMET, Ph.D. Oklahoma 1984. An economic analysis of a Markavian depletion model for petroleum resource.

DAVID K. SMITH, Ph.D. California-Davis 1983. An economic analysis of California egg supply and wholesale-retail price adjustments.

JOHN R. SNYDER, Ph.D. Colorado State 1984. The economics of snow management in the northern Great Plains.

JOHN M. STAATZ, Ph.D. Michigan State 1984. A theoretical perspective on the behavior of farmers' cooperatives.

SPIRO STEFANOU, Ph.D. California-Davis 1983. Impact of scouting information on pesticide application decisions: Cotton in the San Joaquin Valley of California and *lygus hesperus* (Knight).

ABEBAYEHU TEGENE, Ph.D. Iowa State 1983. Rational expectations approach to the modelling of agricultural supply: A case study of Iowa.

R. THAMODARAN, Ph.D. Iowa State 1983. An econometric model for analyzing regional source and use of funds in U.S. agriculture: An application of random coefficients technique.

RANDALL S. THOMAS-PETERHANS, Ph.D. Michigan 1983. The stratification of the production and marketing of livestock in southeastern Niger in the Department of Zinder.

SARAHELEN R. THOMPSON, Ph.D. Stanford 1984. Price performance in thin markets.

RICHARD M. TODD, Ph.D. Minnesota 1983. A dynamic equilibrium model of seasonal and cyclical fluctuations in the corn-soybean-hog sector.

RALPH E. TOWNSEND, Ph.D. Wisconsin-Madison 1983. Some economic issues in the management of Maine's shellfish resources.

GUARACY VIEIRA, Ph.D. Ohio State 1984. Minimum family income and associated farm size in the municipio of Formiga, Minas Gerais, Brazil.

ERIC J. WAILES, Ph.D. Michigan State 1983. A simulation model of U.S. agriculture.

DAVID F. WEIMAN, Ph.D. Stanford 1983. Petty commodity production in the cotton South: Upcountry farmers in the Georgia cotton economy, 1840–80.

ERNST WEISENBOCK, Ph.D. Washington State 1984. A multiperiod linear programming model for planning agricultural production and food supply under emergency conditions.

AREE WIBOONPONGSE, Ph.D. Illinois 1983. Economies of size in multiple cropping in northern Thailand: A random coefficient model approach.

MICHAEL D. WILLIAMS, Ph.D. Colorado-Boulder 1983. An evaluation of alternative uses of Alaskan natural gas.

KELLY D. ZERING, Ph.D. California-Davis 1984. The utility of all-risk crop insurance in the Imperial Valley of California.

CARL R. ZULAUF, Ph.D. Stanford 1983. An analysis of the accumulation of public stocks by a loan program and of the displacement of private storage by public storage.

DECIO ZYLBERSZTAJN, Ph.D. North Carolina State 1983. Price distortions in Brazilian agriculture: An application of duality theory and flexible functional forms.

Manpower, Labor, and Population, including Trade Unions and Collective Bargaining

ANTHONY ABOWD, Ph.D. Chicago 1984. Pay comparisons between federal and nonfederal employees and between blacks and Hispanics using selectivity adjustments and reverse regression.

SEYED-MOHMOUD AGHAZADEH, Ph.D. Nebraska-Lincoln 1983. An econometric model of the demand for, supply of, and the distribution of dentists in Nebraska.

JAHANGIR ALAM, Ph.D. Carleton 1984. Employment performance and cost of labor adjustment in the automobile industry.

RANDY P. ALBELDA, Ph.D. Massachusetts 1983. Black and white women workers in the post-World War II period.

ROBIN L. ALLEN, Ph.D. Northwestern 1984. Physician incentive systems in individual practice association HMOs.

NABEEL A. AL-SALAM, Ph.D. California-Los Angeles 1984. The effects of changes in the age distribution of the workforce on college enrollments and earning-experience profiles.

ALICE AUDI-FIGUEROA, Ph.D. Wisconsin-Madison 1983. Sex, power, and wages: An economic study of comparable worth.

BAHMAN BAHRAMI-DARZAM, Ph.D. Nebraska-Lincoln 1983. Determinants of labor market behavior of registered nurses: The case of Nebraska.

THOMAS R. BAILEY, Ph.D. Massachusetts Institute of Technology 1983. Labor market competition and economic mobility in low wage employment: A case study of immigrants in the restaurant industry.

JACK J. BAROUDI, Ph.D. New York 1984. Job satisfaction, commitment, and turnover among information systems development personnel: An empirical study.

GLORIA J. BAZZOLI, Ph.D. Cornell 1983. The early retirement decision: The influence of health, pensions, and Social Security.

RICHARD S. BELOUS, Ph.D. George Washington 1984. Decentralized planning and employment and training programs: Theories, strategies, and evidence.

ADRIENNE BIRECREE (LYON), Ph.D. Notre Dame 1984. Decertification: A case in point.

DAN A. BLACK, Ph.D. Purdue 1983. Essays on dynamic labor economics.

HELGE BRUNBORG, Ph.D. Michigan 1983. An economic model of fertility, sex, and contraception.

BRUCE S. BUCHANAN, Ph.D. Columbia 1983. Design issues in discrimination and reference testing.

LYNN C. BURBRIDGE, Ph.D. Stanford 1984. The location and mobility of black workers in the United States: An analysis of industries and occupations, 1950–70.

CHARLES A. CALHOUN, Ph.D. Michigan 1983. Lifetime fertility objectives and economic analysis of the age pattern of childbearing.

HIROYUKI CHUMA, Ph.D. State University of New York-Buffalo 1984. Employment contract and specific human capital.

SUZANNE H. CLAIN, Ph.D. Princeton 1984. Intertemporal labor supply decisions of the aging male worker.

GRANT CLOWERY, Ph.D. Chicago 1983. The supply of and demand for accountants: An empirical analysis.

JOSE H. CORTÉS FREGOSA, Ph.D. Texas-Austin 1983. The human capital paradigm and earnings variance: A case study of Guadalajara, Jalisco.

JEREMIAH P. COTTON, Ph.D. Michigan 1983. A comparative analysis of the determinants of black-white and Mexican-American-white male wage differentials.

KENNETH G. DAU-SCHMIDT, Ph.D. Michigan 1984. The effect of consumption commitments on labor supply.

DOUGLAS D. DAVIS, Ph.D. Indiana-Bloomington 1984. Small sample biases and discrimination in employment.

DONALD R. DEERE, JR., Ph.D. Massachusetts Institute of Technology 1983. Labor contracts, unemployment insurance, and efficiency.

ADRIAAN DIERX, Ph.D. Pittsburgh 1984. The effect of region-specific human capital on migration.

ROBERT W. DRAGO, Ph.D. Massachusetts 1983. Success and failure in quality of work life programs, recent development in the United States.

GINETTE DUSSAULT, Ph.D. McGill 1983. Discrimination in the labor market. The case of Montreal office employees.

ELLEN K. ELIASON, Ph.D. Stanford 1983. A stock adjustment of model of contraceptive use in the Philippines.

ROBERT A. FIGLER, Ph.D. West Virginia 1984. Work interruption and human resource management in underground coal mining: An exploratory study.

JEAN W. FLETCHER, Ph.D. Washington-St. Louis 1983. The female wage structure and occupational choice.

DAVID K. FORREST, Ph.D. Western Ontario 1983. Age-structure and unemployment: Some consequences of the postwar baby boom.

HELEN GARRISON, Ph.D. Stanford 1984. Contraceptive effectiveness in Mexico.

SURENDRA K. GERA, Ph.D. Carleton 1984. The efficiency basis and employment impacts of direct job creation in the private sector.

GORDON W. GREEN, JR., Ph.D. George Washington 1984. Wage differentials for job entrants, by race and sex.

DANIEL A. HAGEN, Ph.D. California-Berkeley 1983. Labor market disequilibrium and manpower policy in Sweden.

HANNAH J. HILES, Ph.D. Pennsylvania 1984. Determinants of retirement timing.

CAROL G. HORTON-MANN, Ph.D. Washington State 1984. The relative wages of men and women over business cycles and over time.

SAIFUL HUQ, Ph.D. Boston 1984. Patterns of distribution of rural employment opportunities: A comparative study of two Bangladesh villages.

GEORGE H. JAKUBSON, Ph.D. Wisconsin-Madison 1983. Unobserved heterogeneity in labor economics: Applications of panel data techniques.

ANN M. JOHN, Ph.D. Princeton 1984. The slave population of nineteenth-century Trinidad: A demographic analysis.

ROBERT B. JONES, Ph.D. Georgia 1983. The effect of Social Security on the labor supply of older Americans.

SHULAMIT B. KAHN, Ph.D. Massachusetts Institute of Technology 1983. Occupational safety and worker preferences.

KIPHAM KAN, Ph.D. Southern California 1983. An economic model of the security motive in the demand for children: An application to Malaysia.

ODYSSEUS KATSAITIS, Ph.D. British Columbia 1983. A life cycle model of labor supply.

ETHIOPIA KELETA, Ph.D. Rice 1984. Employment and efficiency of production of small-scale enterprises.

SHERRIE A. KOSSAUDJI, Ph.D. Michigan 1984. Language and labor markets: The effect of English language ability on the labor market outcomes of immigrants.

DAVID A. LAM, Ph.D. California-Berkeley 1983. Economics of household composition in developing countries.

MICHAEL V. LEONESIO, Ph.D. Cornell 1983. In-kind transfers and labor supply.

ELIAS LINKS, Ph.D. State University of New York-Binghamton 1984. An analysis of income differentials in the Republic of South Africa.

JAMES C. LUIZER, Ph.D. Lehigh 1984. The effect of concentration on teachers' salaries in local teacher labor markets.

ELIZABETH A. MCDERMED, Ph.D. North Carolina State 1983. An economic analysis of the effect of mortality risk and health on labor force retirement.

JAMES MCGIBANY, Ph.D. Michigan State 1983. An econometric analysis of the stabilization effective-

ness of the unemployment insurance program.

WALTER S. MCMANUS, Ph.D. California-Los Angeles 1983. Effects of language characteristics on earnings: Hispanic men in the United States.

TIMOTHY J. MALONEY, Ph.D. Wisconsin-Madison 1983. The cyclical labor supply response of married women and disequilibrium unemployment.

TOM MEANS, Ph.D. California-Los Angeles 1983. Compensating wage differentials for income and leisure variation.

TOM C. MOORE, Ph.D. Georgia 1984. Management opposition to unions' impact on certification elections, decertification elections, and state employee bargaining laws.

A. MAUREE O'BRIEN, Ph.D. West Virginia 1983. The labor supply of married women under debt limitations.

DENNIS A. O'CONNOR, Ph.D. Notre Dame 1984. Quality circles: A new direction in labor relations? The case of the ABC Rubber Products Company.

PAUL M. ONG, Ph.D. California-Berkeley 1983. Unions and technological efficiency in auto and steel.

ANTHONY OWUSU-GYAPONG, Ph.D. Queen's 1983. The determinants of strike activity in Canadian manufacturing: An interindustry analysis using panel data.

JAN HWANG PARK, Ph.D. Missouri-Columbia 1984. Studies on cyclical behavior in the labor market.

NORRIS A. PETERSON, Ph.D. Minnesota 1983. Self-selection and the rate of return to education.

MICHAEL J. PILL, Ph.D. Massachusetts 1983. Labor law and economics: Case study of the National Labor Relations Act, as amended.

VUTHIPHONG PRIEBJRIVAT, Ph.D. Chicago 1984. Wage determination of the public sector in Thailand.

WILLIAM C. RANDOLPH, Ph.D. State University of New York-Stony Brook 1983. Employment relationships: Till death do us part?

DAVID L. RYAN, Ph.D. British Columbia 1983. An econometric analysis of the effects of labor market rationing on household labor supply.

SAMUEL A. SHRAGER, Ph.D. Temple 1984. Implicit labor contracts: A theory of specific human capital investment.

JASON K. THOMPSON, Ph.D. Massachusetts 1983. Three components of earnings distribution inequality in metropolitan areas: Industry mix, intra-industry inequality, and inter-industry inequality.

ALI RASSULI, Ph.D. Nebraska-Lincoln 1982. The economic impact of occupational licensure on audiologists.

WENDY L. RAYACK, Ph.D. Wisconsin-Madison 1984. The distributional impacts of recessions: An analysis of earnings-sensitivity by long-run family income class.

JAMES C. ROBINSON, Ph.D. California-Berkeley 1984. Work and health: An economic and policy analysis.

MONIER I. SALEH, Ph. D. Alabama 1983. Organizational effectiveness with the implication for CETA Program: A nationwide study of the program in fiscal year 1980.

GEE SAN, Ph.D. Cornell 1984. The early labor force experience of college students and their postcollege success.

MOHAMMED SHARIF, Ph.D. Boston 1984. Labor supply behavior of the poor in labor abundant LDCs.

PAULA SMITH, Ph.D. Oklahoma State 1984. Occupational attainment and earnings: Differences between aspirants with sex-traditional and sex-nontraditional occupational objectives.

GARY R. SOLON, Ph.D. Princeton 1983. Three studies in labor economics and econometrics.

ELAINE J. SORENSEN, Ph.D. California-Berkeley 1983. Wage discrimination against female-dominated occupations: Theory and evidence.

HARUO TAKAGI, D.B.A. Harvard 1984. Integration between the individual and the organization: An exploratory study of career development under lifetime employment.

PAUL D. THISTLE, Ph.D. Texas A&M 1983. Essays in labor supply.

WILLIAM L. WASCHER, Ph.D. Pennsylvania 1983. Leveling peaks and troughs: Three essays on cyclical and behavioral interaction in the labor market.

Welfare Programs; Consumer Economics; Urban and Regional Economics

CHARLES Z. AKI, Ph.D. Ohio State 1983. Effects of local tax and expenditure policies on the structure and growth of employment in nonmetropolitan counties of the north central United States.

ADELAIDA P. ALICBUSAN, Ph.D. Ohio State 1983. Cash flow analysis of rainfed and irrigated households in Khon Kaen Province, northeast Thailand.

JOAN G. ANDERSON, Ph.D. California-Davis 1984. Indoor temperature and insulation choice: Theoretical and econometric models for policy analysis.

KOFI K. APRAKU, Ph.D. Ohio State 1983. Economic impact of the Lake Erie fishery and other lake industries: An input-output model of the northern Ohio regional economy.

SAID ATRI, Ph.D. State University of New York-Albany 1983. A simultaneous equations model for the demand for hospital care with special reference to quality.

PANAYOTIS P. BASIOTIS, Ph.D. Missouri-Columbia 1983. Food costs, nutrient availability, and nutrient intake of food stamp program eligible households: Analysis in a household production framework.

MICHEL BASTIEN, Ph.D. McGill 1983. Monetary value of permanent changes in health status: An estimate based on individual social welfare functions.

RALPH D. BJONBACK, Ph.D. Colorado State 1984. The impact of the national flood insurance program on

residential property markets.

AXEL H. BOERSCH-SUPAN, Ph.D. Massachusetts Institute of Technology 1984. The demand for housing in the United States and West Germany: A discrete choice analysis.

ALAN S. CANIGLIA, Ph.D. Virginia 1983. Intertemporal effects and the analysis of public housing programs.

NARIS CHAIYASOOT, Ph.D. Hawaii-Manoa 1983. A macroeconomic assessment of the costs of educational policies in Thailand: An elaboration of the Suits-Mason model.

MUHAMMAD A. CHAUDRY, Ph.D. Alberta 1983. Capitalization of local property taxes and public services with applications to efficiency and equity issues.

DONG HOON CHYN, Ph.D. Pennsylvania 1983. Urban housing price determination.

JUDITH K. COHEN, Ph.D. Colorado-Boulder 1984. Estimating the technology of elementary education.

PATRICIA A. COWPER, Ph.D. California-Davis 1984. A spatial analysis of primary health care markets in rural areas.

WILLIAM S. CUSTER, Ph.D. Illinois-Urbana 1984. Hospital competition and physician prices.

JOSEPH A. DI MASI, Ph.D. Boston College 1984. The welfare effects of property tax classification in an urban area: A general equilibrium computational approach.

DAVID S. DRANOVE, Ph.D. Stanford 1983. Economic analyses of physician and patient choice: Physician-induced demand and home nursing care.

KARL DUNZ, Ph.D. California-Berkeley 1984. Some results of economies with land: Equilibria and the core of economies with indivisible and divisible land.

SAMAR M. F. EL-MOFTY, Ph.D. Illinois-Urbana 1983. Health and its effect on the household nonmarket production and consumption.

MARIO R. ESQUIVEL, Ph.D. Texas A&M 1983. Consumer expenditures and economic inequality: An econometric analysis.

PAMELA J. FARLEY, Ph.D. Yale 1984. Hospital and ambulatory services used in the treatment of selected illnesses.

JOHN M. FITZGERALD, Ph.D. Wisconsin-Madison 1983. Life insurance demand and bequests in a household with two earners: Model and evidence from a sample of married couples.

ALIZA FLEISCHER-SHAMAY, Ph.D. Wisconsin-Madison 1983. A U.S. household budget allocation model: The effects of household characteristics under censored response models.

ROBERT B. FRIEDLAND, Ph.D. George Washington 1984. Accumulation of net wealth: Effects of the U.S. retirement system.

MARTIN S. GAYNOR, Ph.D. Northwestern 1983. The effect of income distribution method on equilibrium price in medical group practice.

DAVID L. GEORGE, Ph.D. Temple 1984. Economic rationales for constraints on choice with application to Social Security.

ERNEST P. GOSS, Ph.D. Tennessee 1983. Internal effects labor market: Its effect on occupational and geographic mobility.

JOHN H. GREEN, Ph.D. Pennsylvania 1984. The effects of alternative personal tax rate structures on savings and investment.

JULIA L. HANSEN, Ph.D. California-Berkeley 1983. Segregation and racial transition in urban housing markets.

JOSEPH R. HARRINGTON, Ph.D. Missouri-Columbia 1983. The impact of regional industrial concentration on state unemployment.

HUGO M. HERVITZ, Ph.D. Indiana-Bloomington 1983. An inquiry into the determinants of interregional migration within the framework of a multinomial-logit model.

DAVID K. HORNE, Ph.D. Pennsylvania 1983. The distributional impact of housing in an inflationary environment.

L. W. KEITH HORNER, Ph.D. Carleton 1984. Savings incentives and the retirement income system.

FRANZ JESSEN, Ph.D. Virginia Polytechnic Institute 1984. Marginal tax cost shares and government size.

NANCY A. JIANAKOPLOS, Ph.D. Ohio State 1983. Household wealth accumulation during periods of inflation: Some evidence from longitudinal data.

MELVIN N. JOHNSON, D.B.A. Indiana 1983. A discriminant analysis of selected factors affecting migration changes in nonmetropolitan counties in Georgia, 1960–75.

STEVEN C. JOHNSON, Ph.D. State University of New York-Albany 1983. The factors effecting utilization of inpatient care by New York State medicaid clients.

GRETCHEN JORDAN-BLUMBERG, Ph.D. New Mexico 1984. Application and assessment of the contingent valuation method for federal hazardous waste policies in the Washington, D.C. area.

ALAN L. KESSLER, Ph.D. Boston College 1984. A bureaucratic decision-making model as applied to the New York AFDC Program.

WON NYON KIM, Ph.D. Hawaii-Manoa 1984. A system-wide approach to demand analysis for housing characteristics and effects of demographic variables in Honolulu SMSA.

CHRISTOPHER KLISZ, Ph.D. Wayne State 1984. An empirical study of mortgage redlining in Detroit.

KYU SIK LEE, Ph.D. Hawaii-Manoa 1984. Effect of medical insurance on the demand for medical care in Korea.

ROBERT M. LIPP, Ph.D. State University of New York-Stony Brook 1983. Physician-induced demand and the supply of physicians' services.

MOLLY K. MACAULEY, Ph.D. Johns Hopkins 1984. The

economic value of the geostationary orbital arc, with reference to urban spatial models.

WENDY MAX, Ph.D. Colorado-Boulder 1983. Non-industrial private landowner response to taxation.

DAVID L. MENGLE, Ph.D. California-Los Angeles 1983. Rent control and housing quality: An empirical analysis of the effect of second-generation rent controls on the quality of rental housing.

TERRI J. MENKE, Ph.D. Wisconsin-Madison 1984. Economic welfare and urban disamenities, 1940–78.

DAVID F. MERRIMAN, Ph.D. Wisconsin-Madison 1983. The effect of tax and expenditure limitations on the size and allocation of municipal budgets.

DAVID J. MOLINA, Ph.D. Texas A&M 1983. Cost-of-living index theory under endogenous preferences.

A. B. M. WALIUL I. MONDAL, Ph.D. Ohio State 1984. The role of local government tax and expenditure policies and net immigration on the growth of employment: An econometric study of the North Central region of the United States.

NORMA C. MORGAN, Ph.D. Nebraska-Lincoln 1982. Economic analysis of hospice programs in the United States.

ALFRED MORRISSEY, Ph.D. Notre Dame 1984. An econometric analysis of home energy expenditures and need.

GERMANO M. MWABU, Ph.D. Boston 1984. A household's demand for health services: A study of rural health care with emphasis on Kenya.

JOHN A. NYMAN, Ph.D. Wisconsin-Madison 1984. The effects of medicaid reimbursement policy and information costs on the quality of nursing home care under excess demand conditions.

HAFIZ PASHA, Ph.D. Stanford 1984. Seasonal housing markets: A model of pilgrim housing in Makhah.

TIMOTHY H. QUINN, Ph.D. California-Los Angeles 1983. A more general theory of environment policy with an application to the evolution of groundwater law in California.

BIJAN RAFAILZADEH, Ph.D. Michigan 1984. The economics of bribery in less developed countries with special application to Iran.

CHRISTINE K. RANNEY, Ph.D. California-Davis 1983. A study of the interdependent food stamp program participation and food demand decisions.

MOHSEN SAMI, Ph.D. Nebraska-Lincoln 1983. The role and impact of education on economic and political development of Iran.

EDWARD SAUERACKER, Ph.D. City-New York 1984. Adolescent health, 1966–70: The role of income, race, family background, and preventive care.

JERALD A. SCHIFF, Ph.D. Wisconsin-Madison 1984. Charitable contributions of money and time: The role of government policies.

RONALD H. SCHMIDT, Ph.D. Wisconsin-Madison 1983. Social welfare and optimal depletion: An application to natural gas deregulation.

ROBERTA SCHRODER, Ph.D. State University of New York-Stony Brook 1984. An econometric analysis of readmission to mental health facilities.

ROBERT L. SEIDMAN, Ph.D. Johns Hopkins 1984. Illness uncertainty, risk reduction, and the demand for preventive medical care.

SHANKAR P. SHARMA, Ph.D. Hawaii-Manoa 1984. Alternative theories of agricultural contracts and their policy implications for rural welfare.

AMAR N. SIL, Ph.D. Columbia 1984. Valuation of a corporate pension put option and its effect on pension liability.

STEVEN D. SILVER, Ph.D. California-Berkeley 1984. Value systems and household production of social commodities in the demand for housing.

KENNETH S. SLAYSMAN, Ph.D. Pennsylvania State 1983. The willingness to pay for protection: The case of the accident at Three Mile Island.

J. PALMER SMITH, Ph.D. Columbia 1983. The relationship of resources and municipal service provision in New York City, fiscal years 1978 to 1982.

LONNIE STEVANS, Ph.D. Oklahoma State 1983. Urban property crime and the distribution of income.

MARK E. STOVER, Ph.D. Washington-St. Louis 1984. The price elasticity of the supply of single family detached urban housing.

SHEN SHENG PHILLIP TSENG, Ph.D. George Washington 1984. Economics of tenure choice.

JAMES R. WHITE, Ph.D. Columbia 1983. The effects of zoning on urban housing and land markets.

CALLA J. WIEMER, Ph.D. Wisconsin-Madison 1984. An economic analysis of schistosomiasis control programs in the People's Republic of China.

MARY L. WILLIAMS, Ph.D. Temple 1984. A simulation model of an urban housing market.

ROGER C. WILLIAMS, Ph.D. State University of New York-Stony Brook 1983. An economic analysis of mortgage delinquency under severe stagflation.

MILICA ZARKOVIC, Ph.D. Temple 1984. Economic growth, migration and welfare in the agricultural states of of India, 1960–75.

HAMID ZIAI-AHMADI, Ph.D. Claremont 1984. Inflation and cost of homeownership: Who benefits?

NOTICE TO ALL GRADUATE DEPARTMENTS

The December 1985 issue of the *Review* will carry the eighty-second list of doctoral dissertations in political economy in American universities and colleges. The list will give recipients and titles of doctoral degrees conferred during the academic year terminating June 1985. This announcement is an invitation to send us information for the preparation of the list.

By June 30, please send us this information on 3×5 cards, conforming to the style shown below, one card for each individual. Please indicate by a classification number in the right-hand corner the field in which the thesis should be classified. The classification system is that used by the *Journal of Economic Literature* and printed in every issue.

All items and information should be sent to the Assistant Production Editor, *American Economic Review*, Room 8279, Bunche Hall, University of California, Los Angeles, CA 90024.

JEL Classification No. ________

Name: LAST NAME IN CAPS: First Name, Initial

Institution Granting Degree: ____________________

Degree Conferred (Ph.D. or D.B.A.) ____________ Year ____________

Dissertation Title: ____________________

THE AMERICAN ECONOMIC REVIEW

VOLUME LXXIV

BOARD OF EDITORS

George A. Akerlof
Clive D. Bull
Patricia Danzon
Michael R. Darby
Philip E. Graves
Jack Hirshleifer
Meir Kohn
Frederic S. Mishkin
Sherwin Rosen
Richard Schmalensee
Susan Woodward
Leslie Young

MANAGING EDITOR

ROBERT W. CLOWER

ASSOCIATE EDITOR

JOHN G. RILEY

THE AMERICAN ECONOMIC ASSOCIATION

Executive Office: Nashville, Tennessee

Editorial Office: University of California, Los Angeles, California

Copyright 1984

All Rights Reserved

AMERICAN ECONOMIC ASSOCIATION

CONTENTS OF ARTICLES AND SHORTER PAPERS

CONTENTS OF PAPERS AND PROCEEDINGS

CONTRIBUTORS TO ARTICLES AND SHORTER PAPERS

CONTRIBUTORS TO PAPERS AND PROCEEDINGS

ECONOMICS
Martin Bronfenbrenner
Aoyama Gakuin University, Japan
Werner Sichel and **Wayland Gardner**
Both of Western Michigan University
Complete hardcover edition
Two-volume paperback edition:
MACROECONOMICS • MICROECONOMICS

Study Guide by Rose Pfefferbaum,
Mesa Community College
Instructor's Manual • Test Bank • Computerized
Test Bank • Color Transparencies • Computer
Graphics Package • Now available

For adoption consideration, request an
examination package from your regional
Houghton Mifflin office.

Houghton Mifflin
13400 Midway Rd., Dallas, TX 75234
1900 So. Batavia Ave., Geneva, IL 60134
Pennington-Hopewell Rd., Hopewell, NJ 08525
777 California Ave., Palo Alto, CA 94304

BRONFENBRENNER
SICHEL
& GARDNER
THE RATIONAL
CHOICE

Economics and Finance
Scott, Foresman and Company

Economics
Principles of Political Economy
Second Edition
Daniel R. Fusfeld, The University of Michigan

A unique approach to the principles course, emphasizing a historical and institutionalist perspective.

In the Macroeconomic Material:
- Macroeconomic Instability stressed, highlighted by many examples from the Reagan years
- Flexible price macroeconomic model developed in place of the traditional fixed price model

In the Microeconomic Material:
- New topics including industrial policy, the pros and cons of new business mergers, and Third World debt/trade problems
- The role of big business in the modern economy stressed, including pricing policies

Available January 1985. *©1985, 768 pages, hardbound, Instructor's Manual, Study Guide.*

Dollars and Sense
An Introduction to Economics
Fourth Edition
Marilu Hurt McCarty,
Georgia Institute of Technology

A continuing best-seller, this one-term text presents a balanced selection of macro and microeconomic concepts. Most applications-oriented text available. **Available Now.** *©1985, 320 pages, softbound, Study Guide, Instructor's Manual.*

Understanding Futures Markets
Robert W. Kolb, University of Miami

A publishing first, this text explores all the basic types of futures contracts, the workings of the market, and the role of regulation. Minimal mathematical pre-requisites; several helpful mathematical appendices. **Available Now.** *©1985, 330 pages, softbound.*

Urban Economics
Third Edition
Edwin S. Mills, Princeton University
Bruce Hamilton, Johns Hopkins University

Over 50% new material in this major revision, including urban history, regional economics, and a chapter on LDC cities. Virtually all of the urban problems chapters have been rewritten. *©1984, 450 pages, hardbound.*

Macroeconomics, Second Edition
Microeconomics, Second Edition

Michael E. Bradley, University of Maryland

Brief, rigorous, and exceedingly clear, these major revisions present the elements of modern macro and microeconomics and apply analytical concepts to current issues and problems.
Extensive problem sets are featured.

Macroeconomics

- Integrated aggregate-supply/aggregate-demand model
- Emphasis on inflation, unemployment, and their peculiar trade-offs
- Available January 1985. *©1985, 450 pages, softbound, Study Guide, Instructor's Manual*

Microeconomics

- Indifference curves and isocosts/isoquants treatment integrated into the text body
- New chapter on market failure includes strong coverage of topics such as property rights, Coase's theorem, optimum output of government goods, and benefit-cost analysis
- Available January 1985. *©1985, 480 pages, softbound, Study Guide, Instructor's Manual*

Economics
The Options for Dealing with Scarcity

Frank D. Tinari, Seton Hall University

Introductory text for the one-semester or one-quarter nonmajors course presents an innovative three-part organizational structure which has been honored with an award from the Joint Council on Economic Education. **Available January 1985.** *©1985, 496 pages, softbound, Study Guide, Instructor's Manual/Test File, Transparency Masters.*

Modern Labor Economics
Second Edition

Ronald G. Ehrenberg /
Robert S. Smith
Cornell University

The best-selling text in the field is thoroughly updated and expanded, including the most recent research advances. The most comprehensive and modern treatment of neoclassical labor economics on the market. More public policy applications and technical appendices. **Available January 1985.** *©1985, 540 pages, hardbound, Instructor's Manual.*

Environmental and Natural Resource Economics

Thomas H. Tietenberg,
Colby College

Thorough, up-to-date coverage of environmental economics, with a strong public policy and applications orientation. *©1984, 525 pages, hardbound.*

Economics
Second Edition
Ralph T. Byrns / Gerald W. Stone, Jr.
Metropolitan State College and The University of Colorado at Denver

- Completely redesigned for improved pedagogy and visual appeal
- Analytical rigor, intuitive explanations, and lively writing style
- Thoroughly updated with greater number of definitions and real-world applications
- New MACROSIM, macroeconomics simulation for microcomputers

©1984, 976 pages, hardbound, MACROSIM, new edition of Great Ideas for Teaching Economics, revised Test Bank (computerized and available on both tape and diskettes), Study Guide, Instructor's Manual, and Transparency Masters. Also available in separate macro/micro softbound editions.

Statistics for Business and Economics
Heinz Kohler, Amherst College

Intuitive, applications-oriented approach to real-world statistics. Floppy-disk problems set for IBM and Apple computers focuses on the step-by-step setup of challenging problems and includes a unique program that crunches numbers. **Available Now.** ©*1985*, 760 pages, hardbound, Microcomputer Floppy Disk Problems Set, Instructor's Manual, Study Guide.

Labor Economics and Labor Relations
Robert J. Flanagan, Stanford University
Robert S. Smith / Ronald Ehrenberg
Cornell University

After building up the modern economic labor tools step by step, this text discusses industrial relations based on both market incentives and legal parameters. ©*1984*, 638 pages, hardbound.

American Economic History
Jonathan R. T. Hughes,
Northwestern University

Superbly written history of the American economy from 1607 to the present emphasizes institutional, legal, demographic, and structural change over time.
© 1983, 610 pages, hardbound.

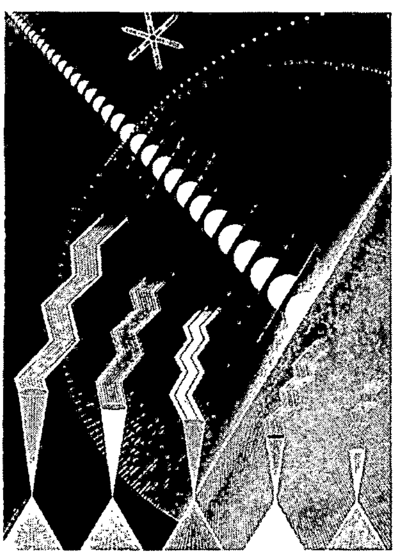
For further information write:
Meredith Hellestrae, Department SA–AER 1900 East Lake Avenue Glenview, Illinois 60025

Scott, Foresman and Company

MACRO-ECO-NOMICS

(Finally, a text as real-world as macroeconomics itself)

Gary Smith, Fletcher Jones Professor of Economics, Pomona College
December 1984, 656 pages, 215 illustrations, 190 graphs
Instructor's Manual by the author



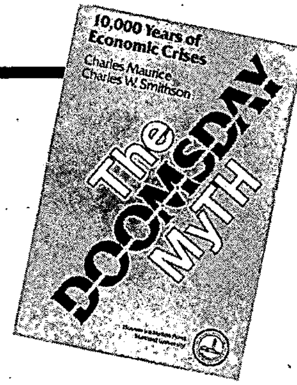
FROM THE FOREWORD

"Students learning macroeconomics are fortunate to have Gary Smith as their guide. At Yale, where he got his Ph.D. in 1971, he was one of the best graduate students I have taught. He stayed at Yale another seven years as a faculty member. We taught together, did research together, and wrote together. We learned from one another. Gary is a first-rate economist and a marvelous teacher of economics. In this book he is sharing his talents with a wider audience. As in the classroom, he will guide you through the controversies . . ., introduce you to the analytical tools of the discipline, relate macroeconomics to economic history and to the real world you live in today. Most important, he will prepare you to be your own macroeconomist and invite you to think for yourself."

JAMES TOBIN
Sterling Professor of Economics, Yale University

Request a copy of MACROECONOMICS today.
On school letterhead, write: Marketing Department, W. H. Freeman and Company, 41 Madison Avenue, New York, N.Y. 10010.

Suddenly the macroeconomics picture has become much clearer.



The Essence of Hayek

edited by Chiaki Nishiyama and Kurt R. Leube
foreword by W. Glenn Campbell

Published on the occasion of Hayek's 85th birthday, this new book is an outstanding collection of his most important and influential contributions. With a biographical portrait of Hayek by Kurt R. Leube, this volume also contains an introduction to Hayek's concepts by Chiaki Nishiyama.

\$27.50, cloth \$13.50, paper August 1984

The Great Depression, 1929-1938
Lessons for the 1980s

by Christian Saint-Etienne

"Ever since the 1930s prophets of doom have been eagerly looking foward to a repetition of the Great Depression. One does not have to be of their number to find lessons in the events leading up to that Depression. We are better able to do this as a result of a new study by a French economist, Christian Saint-Etienne, entitled *The Great Depression, 1929-1938, Lessons for the 1980s.*"

—*Financial Times*
London

\$19.95, cloth \$5.95, paper June 1984

China and Japan
New Economic Diplomacy

by Chae-Jin Lee

The author examines the lively unfolding of Sino-Japanese economic interaction by concentrating on three case studies: a multi-billion dollar steel complex at Baoshan, the joint offshore oil development in the Bohai Sea, and Japanese government loans provided for China's important construction projects.

\$19.95, cloth \$9.95, paper September 1984

The Doomsday Myth
10,000 Years of Economic Crises

by Charles Maurice and Charles W. Smithson

Using nontechnical and often humorous descriptions of ten resource crises that have been faced and solved through the centuries, the authors demonstrate the strengths of a free market.

\$16.95, cloth June 1984

Hoover Institution Press • Dept. 8171
Stanford University • Stanford, CA 94305 • (415) 497-3373

ACADEMIC PRESS

New Titles in Economics from Academic Press . . .

Three Volumes in the ECONOMIC THEORY, ECONOMETRICS, AND MATHEMATICAL ECONOMICS Series . . .

Microeconomic Modeling and Policy Analysis
Studies in Residential Energy Demand

Thomas G. Cowing
Daniel L. McFadden

November 1984, 320 pp., $49.50
ISBN: 0-12-194060-8

CONTENTS: Introduction. Approaches to Energy Demand Modeling. Residential Energy Demand Simulation Models. Residential Energy Policy Analysis: The Alternative Policy Scenarios. Simulation Results from the REEPS Model. Simulation Results from the ORNL Model. A Comparative Analysis of Model Performance: REEPS versus ORNL. Implications. Concluding Remarks. Bibliography. Index.

Asymptotic Theory For Econometricians
Halbert White

1984, 248 pp., $39.00/ISBN: 0-12-746650-9

Devoted solely to large sample theory for economists and econometricians, this work is a collection of results previously scattered throughout the statistics and econometrics literature. Also given are new results that apply to time-series, cross section and longitudinal data for single equations as well as systems of equations. State-of-the-art techniques are introduced and the author provides a comprehensive and unified treatment of statistical estimation and inference for large samples of data. Covering laws of large numbers and central limit theory for independent and dependent data, this work emphasizes the unity of underlying diverse data configurations and associated statistical techniques.

Clear exposition and helpful background for the more difficult material are available along with frequent exercises and selected solutions to provide useful practice and insight. This work serves as an excellent reference and textbook for graduate students in economics, business administration, and statistics.

Stochastic Modelling of Social Processes
Edited by
Andreas Diekmann
Peter Mitter

1984, 352 pp., $30.00/ISBN: 0-12-215490-8

CONTENTS: Introduction. *A. Rapoport,* New Aspects of Stochastic Model Building in the Social Sciences. *M.T. Hannan,* Multistate Demography and Event History Analysis. *A.B. Sørensen,* Interpreting Time Dependency in Career Processes. *A. Diekmann and P. Mitter,* A Comparison of the "Sickle Function" with Alternative Stochastic Models of Divorce Rates. *G.U.H. Seeber,* Semi-Markov and Competing Risks Models with Applications to Occupational Mobility. *J.S. Coleman,* Stochastic Models for Market Structures. *K.C. Land,* Markovian Transition Rates Models of Macro Social Change. *G. Arminger,* Analysis of Event Histories with Generalized Linear Models. *M. Hudec,* Survival Analysis in Heterogeneous Populations—Statistical Models and Concepts. Author Index. Subject Index.

Economic Structure and Performance
Edited by
Moshe Syrquin
Lance Taylor
Larry E. Westphal

December 1984, 608 pp., $79.50
ISBN: 0-12-680060-X

CONTENTS (Sections Headings and Contributors): Structure: *E.S. Mason, G. Ranis, I. Adelman, C.T. Morris, M. Syrquin, S. Bowles, and J. Gurley.* **Strategy:** *A.O. Krueger, S.R. Lewis, Jr., P. Clark, J. Tinbergen, S.W.Y. Kuo, and A. Fishlow.* **Adjustment:** *E.L. Bacha, J. Waelbroeck, J. Behrman, R.K. Sah, B. Balassa, C. Diaz-Alejandro, M. Bruno, and S. Shishido.* **Models:** *A.S. Manne, P.V. Preckel, V. Ginsburgh, S. Robinson, C. Bell, T.N. Srinivasan, R.S. Eckaus, A. Mohie-Eldin, L. Taylor, H. Sarkar, J. Rattso, D. Seers, L.E. Westphal, and J. Cremer.*

Stochastic Models of Control and Economic Dynamics
V. I. Arkin
I. V. Evstigneev
Translated and edited by
E. A. Medova-Dempster
M. A. H. Dempster

February 1985, c. 204 pp., In preparation
ISBN: 0-12-062080-4

CONTENTS: Deterministic Models. The Maximum Principle for Stochastic Models of Optimal Control and Economic Dynamics. Markov Control: The Maximum Principle and Dynamic Programming. Optimal Economic Planning Over an Infinite Horizon: Weak Turnpike Theorems. Approximation of Programmes and Strong Turnpike Theorems. Appendices. References. Index.

Academic Press, Inc.
(Harcourt Brace Jovanovich, Publishers)
Orlando, Florida 32887, U.S.A.
24/28 Oval Road, London NW1 7DX, U.K.

10124

Orlando San Diego New York London Toronto Montreal Sydney Tokyo

For Fastest Service CALL TOLL FREE 1-800-321-5068. To Place An Order From Florida, Hawaii, Or Alaska CALL 1-305-345-4100. In The U.K. CALL (01) 300-0155. Credit Card Orders Only. Send payment with order and save postage and handling. Prices are in U.S. dollars and are subject to change without notice.

AEA sponsored Group Life Insurance for you and your family— at attractive rates!

The AEA Group Life Insurance Plan can help provide valuable supplementary protection—at attractive rates—for eligible members and their dependents.

Because AEA participates in a large Insurance Trust which includes other scientific and technical organizations, the low cost may be even further reduced by premium credits. In the past seven years, insured members received credits on their April 1 semiannual payment notices averaging over 44% of their annual premium contributions. (These credits are based on the amount paid during the previous policy year ending September 30.) Of course future premium credits, and their amounts, cannot be promised or guaranteed.

Now may be a good time for you to re-evaluate your present coverage and look into AEA Life Insurance. Just fill out and return the coupon for more details at no obligation.

Administrator, AEA Group Insurance Program F-8
1255 23rd Street, N.W.
Washington, D.C. 20037

Please send me more information about the AEA Life Insurance Plan.

Name ______________________ Age ______

Address ______________________

City ______________ State ______ Zip ______

Or—call today Toll-Free 800-424-9883
(Washington, DC area, call 296-8030)

Public Enterprise in Mixed Economies: Some Macroeconomic Aspects

This volume contains three papers:

Some Topical Issues Concerning Public Enterprises, by Robert H. Floyd

Toward a Conceptual Framework for Macroeconomic Evaluation of Public Enterprise Performance in Mixed Economies, by Clive S. Gray

The Role of Public Enterprises: An International Statistical Comparison, by R.P. Short.

The three papers should be of interest to government officials, academicians, and others interested in the operations of the public enterprise sector as a whole, or any industrial grouping therein, or even individual enterprises with sufficient weight to have an impact on macroeconomic parameters.

Price: US$12.00, paperbound

Please address orders to:

Publications Unit, Box E–181
International Monetary Fund
700 19th Street, N.W.
Washington, D.C. 20431
Telephone: (202) 473–7430

NOW AVAILABLE
Collected and on Microfiche for the first time.

ADAM SMITH REFERENCES TO THE WEALTH OF NATIONS

Pergamon Press has compiled the complete texts of monographs used by Adam Smith as references for The Wealth of Nations and put them on microfiche. Part of a new Pergamon series, History of Economics, this major collection goes beyond Edwin Cannan's classic Index of Authorities.

Published in two segments, and now available for immediate delivery, this comprehensive collection comprises:

- **Over 2000 Fiche**
- **Over 180,000 Filmed Pages**
- **Two Important Supplementary Collections**
  —46 works from Adam Smith's personal library
  —22 early editions of The Wealth of Nations

For further information write or call:

PERGAMON PRESS
US: Maxwell House, Fairview Park, Elmsford, New York 10523 (914) 592-7700
UK: Headington Hill Hall, Oxford OX3 0BW, England

Please mention The American Economic Review *When Writing to Advertisers*

Harper & Row texts...

New!

Miller
ECONOMICS TODAY ***Fifth Edition***

This new edition remains the clearest and most understandable introductory text on the market. Most books explain economics so that professors can understand it: Miller explains economics so that students can understand it. *ECONOMICS TODAY* retains the outstanding style and superior pedagogy that have made it one of the most successful texts on the market.

December. Available in one volume (832 pages) and separate macro/micro volumes (480 pages each). Instructor's Manual. Test Bank, available on MICROTEST. Transparency Masters. Study Guide.

Bowles/Edwards
UNDERSTANDING CAPITALISM
Competition, Command, and Change in the U.S. Economy
January. 336 pages. Instructor's Manual.

Chow
THE CHINESE ECONOMY
336 pages.

Laidler
THE DEMAND FOR MONEY
Theories, Evidence, and Problems, ***Third Edition***
February. 224 pages.

Melvin
INTERNATIONAL MONEY AND FINANCE
January. 304 pages.

To request examination copies, write to Harper & Row, Suite 3D, 10 East 53d Street, New York, NY 10022. Please include course title, enrollment, and current text.

a tradition of excellence

Miller/Walton/Sexton
ECONOMIC ISSUES IN AMERICA'S PAST
December. 240 pages.

Petersen
BUSINESS AND GOVERNMENT ***Second Edition***
November. 464 pages.

And don't forget...

Lipsey/Steiner/Purvis
ECONOMICS ***Seventh Edition***
1984. 981 pages. Instructor's Manual. Test Bank, available on MICROTEST. Study Guide. Transparency Masters.

Sherman/Evans
MACROECONOMICS
Keynesian, Monetarist, and Marxist Views
1984. 495 pages. Instructor's Manual.

Hammermesh/Rees
THE ECONOMICS OF WORK AND PAY ***Third Edition***
1984. 402 pages. Instructor's Manual.

Waud
ECONOMICS ***Second Edition***
1983. Available in one volume (866 pages) and separate macro/micro volumes (500 and 458 pages, respectively). Study Guide. Instructor's Manual.

Harper & Row

DEFY THE LAWS OF ECONOMETRICS.

Somewhere it is written that large-scale econometric applications require large computers. That the necessary speed and accuracy are available only on a mainframe.

It is also accepted that econometric software must be difficult to learn and complex to use. And that sizeable monthly time-sharing fees are a simple fact of life.

With all due respect to these principles of the past, we proudly present what's new: ESP,™ The Econometric Software Package™ from Alpha Software.® It's a full-scale, fast and highly accurate IBM® PC and XT version of the well-known ESP mainframe program.

And it's already been selected for use by Chase Econometrics, Prudential-Bache and New York Telephone, among others.

Why? Because nothing of the mainframe program is sacrificed. You get the analytical tools you need for sophisticated modeling and forecasting. Nearly all accepted statistical and regression techniques are at your command in one truly powerful program.

In fact, ESP from Alpha is actually enhanced with features like built-in graphics and advanced data management. Plus, ESP interacts like the best personal computer software—with a spontaneous, natural feel that's instinctive, and easy to grasp. It even has on-line help and tutorials.

ESP lets you download from any major data bank including Chase and DRI. Or move your own data from other PC software like 1-2-3,® Multiplan,® dBASE II,® and our own Data Base Manager II.™

But what's best is ESP gives you nearly instant results. Any analysis is performed before your eyes. Models, forecasts, scenarios—as many as you'd like, as often as you'd like, any time you'd like. At no incremental cost.

And you pay just $795 for ESP from Alpha.

For more information and a demonstration version of ESP, send $5.00 to Alpha Software Corporation, 30 B Street, Burlington, MA 01803. Or call us at 1-800-451-1018 (in Massachusetts call 1-800-462-2016).

ESP™
The Econometric Software Package™

ALPHA®
SOFTWARE CORPORATION

ESP and the Econometric Software Package are trademarks licensed to Alpha Software Corp. by MIKROS Corp. IBM is a registered trademark of International Business Machines Corp., 1-2-3 of Lotus, Multiplan of Microsoft, and dBASE II of Ashton-Tate. Data Base Manager II is a trademark of Alpha Software Corp. Alpha Software Corp. is a registered trademark. © 1984 Alpha Software Corp.

Please mention The American Economic Review *When Writing to Advertisers*

Oxford

Spatial Consumer Theory

R.W. BACON, *Lincoln College, Oxford.* By integrating costs of travel, costs of storage and costs of inconvenience into the more conventional models of consumer theory, this book presents a new model which significantly improves economists' ability to plot consumer behavior. The author includes detailed discussions of spatial consumer economics, individual shopping behavior at fixed frequencies, the determination of market areas in the fixed frequency case, the optimal frequency of shopping, and consumer choice over frequency and location.
1984 160 pp. $23.95

Macroeconomics

MICHAEL ARTIS, *University of Manchester.* This compact but comprehensive introduction to macroeconomics takes a non-mathematical approach and assumes only a basic understanding of economic theory. While expounding a Keynesian model, it also allows for an understanding of recent developments such as monetarism and rational expectations.
1984 208 pp.; 68 figs. cloth $29.95 paper $13.50

Energy Planning in Developing Countries

Edited by VLADIMIR BAUM, *Department of Technical Cooperation for Development, United Nations.* These papers focus on the capacity of developing nations to plan and implement the development of indigenous energy supplies. Topics discussed include energy planning and implementation, energy supply and demand management, information and manpower requirements, and the relationship of energy to economic development.
1984 240 pp.; 10 figs. $39.95

Finance and the Firm

EDWARD DAVIS, *University of Aston* and JOHN POINTON, *Plymouth Business School.* This book combines a rigorous treatment of the theory of investment and finance with examples demonstrating the application of the theory in the context of the British financial system. A detailed discussion of the efficiency of alternative methods of investment appraisal is followed by a careful analysis of whether a firm's capital structure and dividend policy are significant determinants of its overall valuation.
1984 304 pp.; 43 figs. cloth $39.95 paper $19.95

Microeconomics

Problems and Solutions

DAVID M. WINCH, *McMaster University.* Course tested over 25 years of teaching, this workbook provides students with an opportunity to apply economic theory to the art of problem solving. The chapters parallel the sequence in which theory is presented in most textbooks and are divided into three parts: Part I contains problems, Part II contains hints, and Part III contains full solutions.
1984 224 pp. paper $9.95

Applied Economics in Banking and Finance

Third Edition

H. CARTER, *Economist with Buckmaster and Moore* and I. PARTINGTON, *Leeds Polytechnic.* This text presents both the background of the complex institutional structure of the U.K. financial system as well as an analysis and evaluation of the current system. Major revisions to this edition bring it completely up to date and keep it abreast of the Institute of Bankers' requirements.
1984 390 pp.; 14 figs. cloth $29.95 paper $13.50

Prices are subject to change.
To order, send check or money order to
Assistant Marketing Manager, Humanities & Social Sciences

Oxford University Press

200 Madison Avenue • New York, New York 10016

A New Vision in Economics

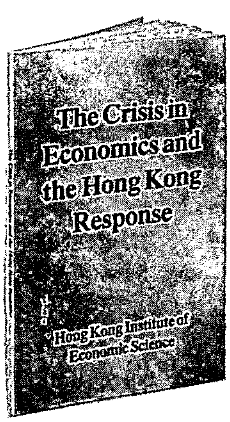
That economics has been in a crisis is commonly recognized. But what is needed to lead economics out of it has yet to emerge. Post-Keynesians, supply-siders, rational expectationists, neo-Austrians, institutionists, evolutionists, liberal and radical economists alike apply the particular strengths of their schools to compensate for particular weaknesses of the orthodoxy. We are still awaiting a systematic alternative capable of confronting fully and squarely the many problems of our present-day economic reality and thereby supplanting the orthodoxy.

Here is the. . .

Hong Kong Solution

to the Crisis in Economics

The following booklet contains an analysis by the Hong Kong Institute of Economic Science of the nature and origin of the crisis in contemporary economics. This analysis is a prelude to the Hong Kong solution. It argues for the need to adopt a new methodology and to develop a more contentful set of first principles for economics. Please fill out and return the coupon for your free copy.



FREE

I'd like to receive, *gratis*, a copy of the booklet ***The Crisis in Economics and the Hong Kong Response***

Name (please print) ______________________

Address ______________________

____

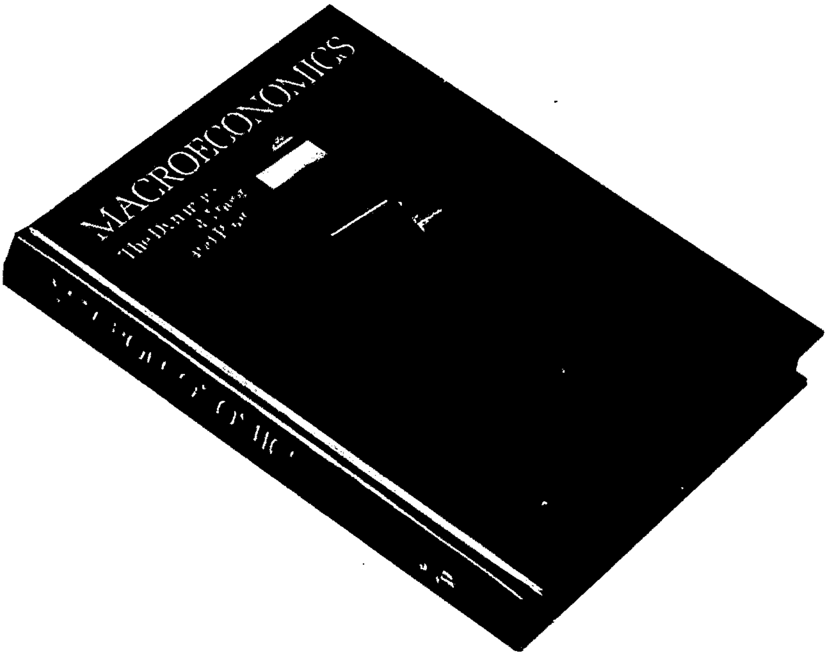
Organisation ______________________

Phone ______________ Date ________

Please send this form to our U.S. correspondence address. Allow 3-4 weeks for free delivery.

Hong Kong Institute of Economic Science
Head Office: 48, Hillwood Road, Suite 1304, Kowloon, Hong Kong Tel: 3-7238841
U.S. Correspondence: 37782 Los Arboles Drive, Fremont, CA 94536 Tel: 415-793-1069

A New Standard Of Excellence



MACROECONOMICS: The Dynamics Of Theory And Policy—an exciting and successful new text by William J. Boyes:

- Provides a clear integration of macro theory and policies within the historical context of their development
- Progresses rapidly to a development of the Aggregate Demand/Aggregate Supply framework and then uses it to unify the treatment of theory
- Devotes an entire chapter to an examination of alternative expectations schemes–including rational and adaptive expectations models
- Applies the unifying AD/AS Model to examine such current issues as tax policy, budgets and deficits, Social Security, declining productivity growth, the political business cycle, and the Third World Debt
- Presents a balanced treatment of current controversies–including rules versus discretion, the explanation of stagflation, contract- and information-based theories of the business cycle, and supply-side economics

For more information on this text and other South-Western titles, contact:

Lew Gossage
COLLEGE DIVISION
SOUTH-WESTERN PUBLISHING CO.
5101 Madison Road
Cincinnati, OH 45227.

BROOKINGS

Energy, Economics, and Foreign Policy in the Soviet Union

Ed A. Hewett

Over the past few years, Western perceptions of Soviet energy prospects have shifted dramatically—from a 1977 CIA prediction of rapidly falling oil output in the 1980s to the current concern that an energy-abundant Soviet Union will flood Western Europe with natural gas. In this major new study, Hewett analyzes the prospects for the Soviet energy sector through the rest of the decade, the implications for Soviet economic performance, and possible links with Soviet foreign policy, particularly in the Persian Gulf area.

1984/228 pages/$10.95 paper
$28.95 cloth

The Australian Economy: A View from the North

Richard E. Caves
and Lawrence B. Krause, editors

Australia's economic performance, like that of other industrial countries, deteriorated significantly between 1973 and 1983. But the sources of Australia's economic problems differ greatly from those of other countries, largely because of its distinctive characteristics—low population, isolation from its trading partners, and abundance of minerals and pastoral land in relation to its land mass. In this volume, nine economists from the United States and Canada examine the economic problems of Australia and bring new insights and perspectives to those problems—a view from the North.

1984/c. 450 pages/$12.95 paper
$32.95 cloth

U.S.-Canadian Economic Relations: Next Steps?

Edward R. Fried and Philip H. Trezise

Trade is a major force tying the United States and Canada together, but its dimensions are not always understood. In this Dialogue on Public Policy, distinguished Canadians and Americans—businessmen and present and former government officials—give their views on the bilateral economic relationship. They emphasize trade and, more specifically, the possibilities for the removal of barriers to its further growth, ranging from new bilateral sectoral free-trade arrangements to a full free-trade area.

1984/c. 150 pages/$9.95 paper

Latin America: The Crisis of Debt and Growth

Thomas O. Enders
and Richard P. Mattione

In this analysis of the structure of the Latin American economies and the external and internal shocks that those countries experienced from 1979 to 1982, the authors examine various policy options for confronting the current crisis. They stress the likely difficulties in restoring both growth and debt-servicing capabilities during the 1980s and the particular importance of expanded trade between Latin America and the industrialized world to any resolution of Latin America's economic difficulties.

1984/66 pages/$6.95 paper

The Brookings Institution

1775 Massachusetts Avenue, N.W.
Washington, D.C. 20036
(202) 797-6258

Options for Tax Reform

Joseph A. Pechman, editor

Public opinion polls indicate that the income tax, once regarded as the fairest tax, is now the least popular in the federal system. The idea of tax reform to simplify the federal income tax has gained momentum from the explosion of the federal deficit and the resulting need to increase federal revenue. In this Dialogue on Public Policy, tax experts examine the current debate over tax policy and provide a detailed evaluation of various tax reform proposals.

1984/c. 150 pages/$9.95 paper

Can America Compete?

Robert Z. Lawrence

Is America deindustrializing? Is the rest of the world out-competing American industry? In this timely examination of the performance of U.S. manufacturing, Lawrence finds that, contrary to recent fears, international trade competition has not induced a deindustrialization of America. Challenging many widely held assumptions about U.S. industrial performance, he evaluates proposals for new U.S. industrial policies and offers suggestions for policies to facilitate structural change.

1984/156 pages/$8.95 paper
$22.95 cloth

Economic Choices 1984

Alice M. Rivlin, editor

Hailed as an important contribution to the coming national debate on economic choices during the 1984 election campaign and in the next Congress, this volume examines the ills of the U.S. economy and prescribes a painful solution. It lays out a practical plan for cutting the federal deficit, increasing flexibility in domestic spending, and reforming the federal tax system. It also suggests ways to cope with structural change in American industry and to reverse recent increases in poverty.

1984/171 pages/$8.95 paper
$22.95 cloth

Tax Incentives and Economic Growth

Barry P. Bosworth

"This new book is a detailed and even-handed survey of what economists know about the impact of taxes on decisions to work, save, and invest—the three activities the Reagan supply-side economic program was supposed to promote. . . . Supply-siders won't care for it very much, but the debate would be advanced if they were to meet the challenge implicit in this volume: provide the missing theory and empirical data to back up the supply-side case." John Berry, *Washington Post.*

1984/220 pages/$9.95 paper
$26.95 cloth

The Brookings Institution

1775 Massachusetts Avenue, N.W.
Washington, D.C. 20036
(202) 797-6258

BROOKINGS

MATHEMATICAL AND STATISTICAL PROGRAMMING PACKAGE FOR YOUR IBM PC

FAST • EASY TO USE • POWERFUL

GAUSS™

YOU'VE NEVER SEEN ANYTHING LIKE IT!

GAUSS is a sophisticated mathematical and statistical programming package for the IBM PC and compatibles. It combines **speed, power, and ease of use** in one amazing program.

GAUSS allows you to do essentially **anything you can do with a mainframe** statistical package — and a lot more.

Personal computers are **friendly, convenient, and inexpensive.** So is **GAUSS. GAUSS** is not just a stripped-down mainframe program. **GAUSS** has been designed from the ground up to take advantage of all of the conveniences of a personal computer. After trying **GAUSS,** you may never use a mainframe again.

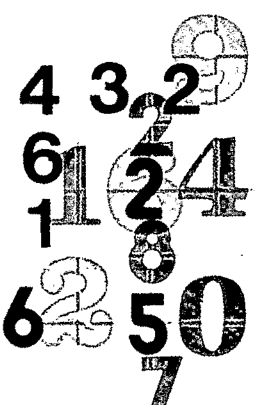
GAUSS comes with programs written in its matrix programming language that allow you to do most **econometric procedures,** including OLS, 2SLS, 3SLS, PROBIT, LOGIT, MAXIMUM LIKELIHOOD, and NON-LINEAR LEAST SQUARES.

In the current version, **GAUSS** will accept up to 90 variables in a regression. There is **no limit** on the number of observations.

GAUSS will do a regression with 10 independent variables and 800 observations in under 4 seconds — and with 50 variables and 10,000 observations in under 18 minutes. It will compute the **maximum likelihood** estimates of a binary logit model, with 10 variables and 1,000 observations, in 1-2 minutes, depending upon the number of iterations required.

GAUSS allows you to do **complicated statistical procedures** that you would never imagine trying on a mainframe. It is **easy to program** almost any routine, and **GAUSS** is so **fast** that it can do almost any job. But the nicest thing of all is that the cost of time on your personal computer is essentially zero!

GAUSS is an **excellent teaching tool.** It makes programming easy and allows students to focus on concepts and techniques.



If you can write it mathematically, you can write it in **GAUSS.** Furthermore, you can write it in **GAUSS** almost exactly the way you would write it mathematically.

GAUSS is **10-15 times faster** than other programs that use the 8087, and **15-100 times faster** than other programs that do not use the 8087.

As in APL, single statements in **GAUSS** can accomplish what might take dozens of lines in another language. However, **GAUSS** provides you with additional **powerful numerical operators and functions** — especially for statistics and the solution of linear equations — that are not found in APL. And, of course, the syntax in **GAUSS** is much more natural (for most of us) than that in APL.

GAUSS has **state-of-the-art numerical routines** and **random number generators.**

GAUSS is **extremely accurate.** It allows you to do an entire regression in 19 digit accuracy. It will compute the Longley benchmark coeficients in 5 hundredths of a second with an average of 11 correct digits! (Try that on a mainframe!)

GAUSS, with its built-in random number generators and powerful functions and operators, is an excellent tool for doing **simulations.**

GAUSS and the 8087 NUMERIC DATA PROCESSOR GIVE YOU MINICOMPUTER PERFORMANCE ON YOUR DESKTOP.

SPECIAL INTRODUCTORY OFFER
With 30 Day Money Back Guarantee .... Reg. 395.00 **$250.00**

GAUSS requires an IBM PC with at least 256K RAM, an 8087 NDP, 1 DS/DD disk drive, DOS 2.0 (or above).

IBM is trademark of IBM Corporation.

Call or Write **APPLIED TECHNICAL SYSTEMS**
P.O. Box 6487, Kent, WA 98064
(206) 631-6679

Come and get a demonstration of GAUSS at our booth in the exhibitor's area at the AEA meetings in Dallas, December 28-30, 1984.

NORTH-HOLLAND ANNOUNCES

NEW BOOKS • NEW BOOKS • NEW BOOKS • NEW BOOKS • NEW

Regional and Industrial Development Theories, Models and Empirical Evidence

edited by ÅKE E. ANDERSSON, WALTER ISARD *and* TÖNU PUU

STUDIES IN REGIONAL SCIENCE AND URBAN ECONOMICS, Vol. 11

1984. xviii + 405 pages.
US $57.75/Dfl. 150.00. ISBN 0-444-87595-6

This proceedings volume consists of 30 selected papers by regional scientists from both East and West, in memory of the great Swedish theorist, Tord F. Palander. It clearly testifies to the admirable progress that has been made in the development of static and comparative static analysis of spatial systems. The work also pays close attention to dynamic analysis of spatial systems, an approach which hitherto has developed to a lesser degree. Although the main emphasis is on the dynamics of spatial economic structure, the selection of papers represent a full state-of-the-art picture of regional science.

Economic-Environmental Modeling in a National-Regional System

An Operational Approach with Multi-Layer Projection

by WILHELMUS A. HAFKAMP

STUDIES IN REGIONAL SCIENCE AND URBAN ECONOMICS, Vol. 10

1984. xvi + 236 pages.
US $44.25/Dfl. 115.00. ISBN 0-444-87592-1

This study deals with economic-environmental modeling for the analysis of interrelated effects of environmental and economic policy on the economic system and environmental quality. Recognizing the communication difficulties inherent to integrated economic-environmental models, a new design procedure for economic-environmental models is proposed.
The principle of multi-layer projection is proposed to improve the manageability of large, integrated economic-environmental models. Multiobjective analysis is used as a primary means for the analysis of trade-offs and conflicts between interest groups in economic-environmental decisonmaking.

Essays in the Economics of Exhaustible Resources

Edited by M.C. KEMP *and* N. LONG

CONTRIBUTIONS TO ECONOMIC ANALYSIS, VOLUME 150

1984. x + 250 pages
US $57.75/Dfl. 150.00
ISBN 0-444-86791-0

The essays in this volume reconstruct substantial parts of economic theory in the light of resource exhaustibility. Among the important topics examined are: the optimality of competitive markets, especially when resource stocks are of unknown extent; the dynamic economic theory of oligopoly; the problem of survival in closed and open economies; and the optimal order of extraction of several heterogeneous deposits. Each section of the book contains a detailed survey of recent developments.

The Performance of Public Enterprises

Concepts and Measurement

Edited by MAURICE MARCHAND, PIERRE PESTIEAU *and* HENRY TULKENS

STUDIES IN MATHEMATICAL AND MANAGERIAL ECONOMICS, Volume 33

1984. x + 296 pages.
US $42.25/Dfl. 110.00. ISBN 0-444-87551-4

This book contains new contributions on the normative, positive and empirical issues of public enterprise behavior. An introductory chapter deals with the performance concept from those three points of view. Four chapters deal with optimal decisions relating to pricing and uncertainty. Three further chapters deal with controllability and principal-agent relations in public enterprises. The last three chapters are empirical studies on public performance in mixed insurance markets, postal activity and electricity regulation.

ELSEVIER SCIENCE PUBLISHERS
P.O. Box 211, 1000 AE Amsterdam, The Netherlands

For customers in the U.S.A. and Canada:
ELSEVIER SCIENCE PUBLISHING CO., INC.
P.O. Box 1663, Grand Central Station, New York, N.Y. 10163

0892NH/BKS/ECON

The eighth—and election year—edition of the highly successful reader!

"An excellent book of supplemental readings for an introductory macroeconomics course. The articles are more current and politically diverse than most readings in the field."
—AAAS Science Books & Films

INTRODUCTORY MACROECONOMICS 1984–85

Readings on Contemporary Issues

Edited by PETER D. McCLELLAND

The eighth in a series of highly successful and widely adopted annual readers for all introductory economics courses, *Introductory Macroeconomics 1984–85* contains more than fifty articles by leading economists published in periodicals as recently as May 1984. The articles are chosen to reflect the entire spectrum of economic opinion, from left-wing to right-wing, from liberal to conservative, from Keynesian to monetarist to supply-side advocate. Writers featured in this past year's reader included well-known commentators on macroeconomic problems such as Kenneth Arrow, Barbara R. Bergmann, Karl Brunner, Milton Friedman, Henry A. Kissinger, Joseph A. Pechman, Paul Craig Roberts, Paul Samuelson, David A. Stockman, and Paul A. Volcker.

The provocative articles and essays are selected from a diverse range of popular and scholarly sources, including *The Wall Street Journal, The Economist, Time, The Progressive, Dollars & Sense, New Republic, Economic Outlook USA, The New York Times, Challenge, U.S. News & World Report, Society, Newsweek,* Federal Reserve publications, and other major magazines and journals. The issues covered in *Introductory Macroeconomics 1984–85* include: inflation and unemployment, monetary and fiscal policy, institutional changes in money and banking, the growing federal deficit and proposals to reduce it, tax-based incomes policy, economic growth and reindustrialization, and poverty, welfare, and income distribution. $9.95 softcover

Send for a free examination copy today

CORNELL UNIVERSITY PRESS
P.O. Box 250, Ithaca, New York 14851

Please send one copy of *Introductory Macroeconomics 1984–85* edited by Peter D. McClelland

I will be considering it for adoption in my course ____________ which has an expected enrollment of ________. The text I am now using is ____________.

Name ____________

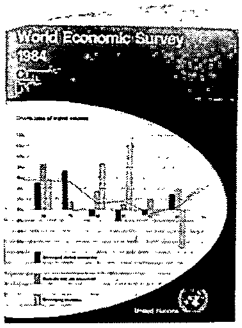
Institution ____________

Department ____________

Street ____________

City ____________ State ________ Zip ________

PUBLICATIONS FROM THE UNITED NATIONS



Catalogue of our International Publications available free upon request

UNITED NATIONS PUBLICATIONS
Room DC2-853,
New York, N.Y. 10017
Palais des Nations,
1211 Geneva 10,
Switzerland

ECONOMIC PUBLICATIONS

WORLD ECONOMIC SURVEY 1984
Current trends in the World Economy. Gives the growth in the world economy and current policy stances, international trade and payments, adjustment and international capital flows to developing countries.
E.84.II.C.1 **$12.00**

ECONOMIC AND SOCIAL SURVEY OF ASIA AND THE PACIFIC
Contains the survey of world economy, major constraints on regional progress, the economic and social survey of South Asia, South-east Asia, South Pacific island countries, China, Iran, Hong Kong, the Republic of Korea and Japan.
E.84.II.F.1 **$18.00**

ECONOMIC SURVEY OF LATIN AMERICA
This publication views trends in the Latin American economy such as inflation, trade, output, external finance and foreign trade, plus the economic evolution of the individual countries.
E.83.II.G.2 **$42.00**

ECONOMIC SURVEY OF EUROPE
Issued annually since 1948
The Survey is an internationally recognised economic report. It ranks among the most highly regarded analyses of economic trends and provides an outlook for the future. Covering a wide range of research, it records the changing patterns of trade and output in Europe and North America.
E.84.II.E.1 **$35.00**

New from RFF

EMISSIONS TRADING
An Exercise in Reforming Pollution Policy
T. H. Tietenberg

1984 240 pages $22.50 hardcover

ENERGY, FORESIGHT, AND STRATEGY
Thomas J. Sargent, editor

Jan. 1985 288 pages $18.00 paperback

MEASURING THE BENEFITS OF CLEAN AIR AND WATER
Allen V. Kneese

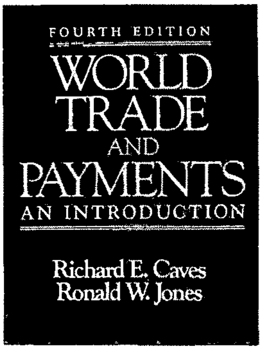
1984 172 pages $5.95 paperback

Order directly from Resources for the Future,
P.O. Box 4852, Hampden Station,
Baltimore, MD 21211

Resources for the Future

Please mention THE AMERICAN ECONOMIC REVIEW *When Writing to Advertisers*

Texts you can count on...



New edition!

WORLD TRADE AND PAYMENTS

An Introduction

Fourth Edition

Richard E. Caves and Ronald W. Jones

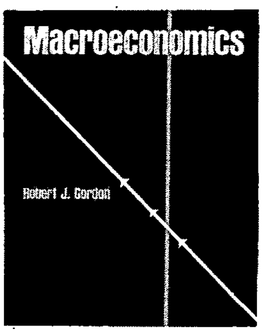
Thoroughly updated, this revision of the leading text for international economics features a reworked,up-to-date treatment of trade theory and strengthened coverage of international finance. And this clear, modern account is now more useful than ever — with supportive chapter-ending exercises and problems new to this edition.
Available Now! #132276/cloth/544 pages

New!

MANAGERIAL ECONOMICS

Robert F. Rooney

This comprehensive new text is more than merely a price theory text disguised as a managerial economics text. Rooney's fresh approach presents microeconomics in a more useful way for business students. The text features worked-out cases that show students how theory can be applied to practical business decisions and problems. Calculus is used where appropriate, and numerous problems are included at the end of each chapter.
Available in February! #755966/cloth/480 pages/with Instructor's Manual



Highly successful!

MACROECONOMICS

Third Edition

Robert J. Gordon

Users across the nation agree that Gordon's Third Edition is the best text — with the best package — for the intermediate macro course. Simplified presentation of inflation and unemployment as well as the latest data available enhance this widely-used text's superb integration of modern theory with numerous real-world cases.
#321079/cloth/720 pages/1984/with Instructor's Manual, Student Workbook, and *The Gordon Update* newsletter

MICROECONOMIC THEORY AND APPLICATIONS
Edgar K. Browning and Jacquelene M. Browning
#112232/1983/cloth
604 pages/with Instructor's Manual and Study Guide

PUBLIC SECTOR ECONOMICS
Second Edition
Robin W. Boadway and David E. Wildasin
#100528/1984/cloth
480 pages

ENVIRONMENTAL ECONOMICS AND POLICY
Paul B. Downing
#191809/1984/cloth
350 pages

from Little, Brown.

College Division • 34 Beacon Street • Boston, Massachusetts 02106

Please mention THE AMERICAN ECONOMIC REVIEW *When Writing to Advertisers*

New Titles from the

National Bureau of Economic Research

NBER

A Retrospective on the Classical Gold Standard, 1821-1931
Edited by **Michael D. Bordo** *and* **Anna J. Schwartz**
Cloth $65.00 694 pages

The Structure and Evolution of Recent U.S. Trade Policy
Edited by **Robert E. Baldwin** *and* **Anne O. Krueger**
Cloth $50.00 448 pages

R & D, Patents, and Productivity
Edited by **Zvi Griliches**
Cloth $50.00 524 pages

The International Transmission of Inflation
Michael R. Darby *and* **James R. Lothian**
With the collaboration of **Arthur E. Gandolfi, Anna J. Schwartz,** *and* **Alan C. Stockman**
Paper $22.95 (est.) 744 pages

Exchange Rate Theory and Practice
Edited by **John F. O. Bilson** *and* **Richard E. Marston**
Cloth $58.00 544 pages

Economic Transfers in the United States
Edited by **Marilyn Moon**
Cloth $42.00 398 pages

The Taxation of Income from Capital
A Comparative Study of the United States, the United Kingdom, Sweden, and West Germany
Edited by **Mervyn A. King** *and* **Don Fullerton**
Cloth $35.00 360 pages

Exchange Rates and International Macroeconomics
Edited by **Jacob A. Frenkel**
Cloth $43.00 392 pages

Published by

CHICAGO

THE UNIVERSITY OF CHICAGO PRESS
5801 South Ellis Avenue, Chicago, IL 60637

Please mention THE AMERICAN ECONOMIC REVIEW *When Writing to Advertisers*

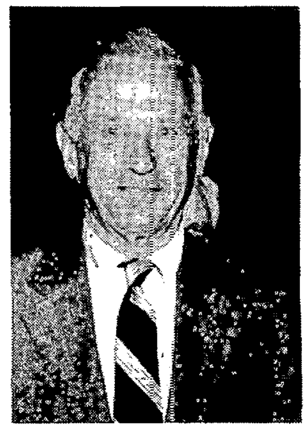
Trade Liberalization among Major World Trading Areas

John Whalley

This book provides a much needed quantitative response to the classic question of who gains and who loses in trade liberalization and shows how important the process is for the global economy. It contributes significantly to the debate concerning trade between developed and developing countries.
$30.00

Multinational Excursions

Charles P. Kindleberger



Multinational Excursions brings together Kindleberger's lucid and often outspoken pieces prepared over the past decade and covering all aspects of the role of multinational corporations in the international economy.

"This collection of articles, reviews, and testimony is more evidence that Charlie Kindleberger is one of the most stimulating applied economists of his generation. There are occasions when the reader disagrees with the insights and conclusions, but rarely, if ever, is the reader bored."–Robert C. Aliber, University of Chicago
$22.50

Deregulation and the New Airline Entrepreneurs

John R. Meyer and Clinton V. Oster, Jr.

"This is the book we have been waiting for; Meyer and Oster have brought us an orderly and comprehensive look at the chaotic events since airline deregulation. It is must reading for the industry and students of the effects of deregulation."
–D. Daryl Wyckoff, James J. Hill Professor of Transportation, Harvard University. Regulation of Economic Activities Series, edited by Richard Schmalensee
$22.50

The Cost of Capital

Estimating the Rate of Return for Public Utilities

A. Lawrence Kolbe and James Read, with George R. Hall

This book evaluates the five major estimation methods used to estimate the required rate of return to investors, providing a systematic assessment of the strengths of each and surveying new methods as well. It is a thorough review and critique that will prove valuable to all members of the regulatory community. A Charles River Associates study.
$25.00

Sources of International Comparative Advantage

Theory and Evidence

Edward E. Leamer

"Combining theory ingeniously with econometrics to reopen the question of the validity of the Heckscher-Ohlin theory of comparative advantage, this is an important work."–Jagdish N. Bhagwati, Arthur Lehman Professor of Economics, Columbia University
$45.00

The Wicksell Lectures

A Search-Equilibrium Approach to the Micro Foundations of Macroeconomics

Peter A. Diamond

Inaugurated in 1958 by the Swedish Economic Association, The Wicksell Lectures have rapidly become one of the most highly regarded forums of economic theory and policy. This book presents the 1983 lectures given by Peter A. Diamond, Professor of Economics at MIT.
$7.50

The Forces of Economic Growth and Decline

Paolo Sylos-Labini

Innovation, changes in market structure, and changes in income distribution are the forces that drive the general process of economic growth or decline. This is the concept that unifies these essays written between 1954 and 1983 by the noted economist Paolo Sylos-Labini. "Sylos-Labini has a keen eye for what is important at the junction of micro-and macroeconomics. Stressing disequilibrium and dynamics, he provides a perspective neglected by economists of the Chicago and (two) Cambridge schools."—F. M. Scherer, Professor of Economics, Swarthmore College
$27.50

A Game-Theoretic Approach to Political Economy

Volume 2 of *Game Theory in the Social Sciences*

Martin Shubik

This companion volume to *Game Theory in the Social Sciences: Concepts and Solutions* sketches a unification of several branches of political economy on the basis of the theory of games. Its five parts cover Basics, Oligopoly, Cooperative Models of Closed Economic Systems, Strategic Models of Closed Economic Systems, and Externalities and Public Goods.
$47.50

Now available in paperback

The Origins and Development of Labor Economics

Paul J. McNulty

$7.95

Controlling Hospital Costs

The Role of Government Regulation

Paul L. Joskow

$8.95

28 Carleton Street
Cambridge, MA 02142

THE MIT PRESS

Please mention THE AMERICAN ECONOMIC REVIEW *When Writing to Advertisers*

The Issues are

ECONOMIC

INFLATION AND THE ECONOMIC WELL-BEING OF THE ELDERLY

Robert L. Clark, George L. Maddox, Ronald A. Schrimper, and Daniel A. Sumner

The first in-depth examination by economists of the commonly held assumption that the elderly suffer disproportionate losses of real income in periods of inflation.

The authors provide an economic model of well-being and indicate how individual welfare is affected by price changes. They demonstrate that, although personal income drops with retirement, the indexing of social security and many other transfer programs means that much of the nominal income of older persons rises automatically with changes in consumer prices. A fresh and thorough reassessment of an issue of equal importance to economists, social gerontologists, and policy administrators in social welfare.

$18.50

IMMIGRATION POLICY AND THE AMERICAN LABOR FORCE

Vernon M. Briggs, Jr.

Since the late 1970s and early 1980s, the United States has been in the midst of the largest influx of immigrants in its history. During this period, argues Vernon Briggs, political considerations have dominated contemporary immigration policy, with little concern for their labor market implications.

IMMIGRATION POLICY AND THE AMERICAN LABOR FORCE charts the evolution of U.S. immigration policies toward all types of immigrants—legal immigrants, illegal immigrants, refugees, nonimmigrant workers, border commuters—and analyzes the impact of these policies upon American labor practices. Briggs calls for a comprehensive policy reform and proposes a number of alternative options, including the establishment of immigration ceilings that are flexible and responsive to the nation's employment trends and a return to occupational preferences as the basis for allowing immigrants to enter the country.

$26.50

THE JOHNS HOPKINS UNIVERSITY PRESS

Baltimore, Maryland 21218

UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN — SAUDI ARABIA

The College of Industrial Management will have faculty positions (all ranks) open for the academic year 1985-86, starting September 1, 1985 in: **Economics**

ACADEMIC QUALIFICATIONS AND EXPERIENCE

Applicants must have a Ph.D. in the field and some teaching experience.

The College of Industrial Management is a young growing college, offering curricula equivalent to the standards of the AACSB. Research is encouraged and high quality proposals are awarded university research grants.

Language of instruction is English.

Minimum regular contract for two years, renewable. Competitive salaries and allowances. Air conditioned and furnished housing provided. Free air transportation to and from Dhahran each year. Attractive educational assistance grants for school age dependent children. All earned income without Saudi taxes. Ten months duty each year with two months vacation with salary. There is also possibility of selection for the University's on-going summer program with good additional compensation.

Apply before December 31, 1984, with complete resume on academic, professional and personal data, list of references, publications and research details, and with photocopies of degrees and/or transcripts, including home and office addresses and telephone numbers to:

UNIVERSITY OF PETROLEUM & MINERALS
HOUSTON OFFICE, DEPT. 290
5718 WESTHEIMER, SUITE 1550
HOUSTON, TEXAS 77057

ANNOUNCING A NEW STATISTICAL PACKAGE
FOR YOUR PERSONAL COMPUTER!

- **Ordinary Least Squares Estimation**
- **Polynomially Distributed Lag Estimation**
- **Two Stage Least Squares Estimation**

- **Cochrane-Orcutt estimation feature in each of above**
- **Usual summary statistics and *more* including D.W., autocorrelation coeff., log of likelihood, correlation matrix, var-cov matrix for estimated coefficients, etc.**
- **Prediction with confidence and prediction intervals in OLS**
- **Handles large numbers of variables and observations**

Price: $99.50 US/$129.50 CAN*

Please send me a CAN-AM STATISTICAL PACKAGE. ( ) The payment is enclosed.
( ) Please charge $129.50 CAN to my ( ) VISA or ( ) MasterCard account.

Credit Card # ____________: Expiry date ____________

Name ____________ Signature (if charged) ____________

Address ____________

*Ontario residents: Please add 7% provincial sales tax.

I use ( ) Apple ( ) Commodore 64 ( ) IBM PC ( ) Pet/SuperPet.

To order please mail this coupon (or photocopy). Make cheque or money order payable to:

CAN-AM FINANCIAL CONSULTING
177 Caddy Ave., Sault St. Marie, Ont. P6A 6H7 CANADA

An Economic Theorist's Book of Tales
GEORGE A. AKERLOF
This collection of essays exploring the consequences of making non-standard economic assumptions breaks with traditional economic theory, which relied upon a tacit and "classical" set of assumptions that have gradually acquired a life of their own in terms of how economists write, think, and justify economic models. Cloth $29.95 Paper $8.95

On Appraising the Performance of an Economic System
RUTLEDGE VINING
"Vining's emphasis is on the simple fact that relevant political choices are not among alternative *distributions* or *allocations*, but must be among alternative sets of institutional arrangements that generate patterns of distribution and allocation. If this book can bring economists to recognize this simple fact, discussion in modern political economy will be dramatically improved. I therefore recommend this book to all economists about the very foundations of their discipline."—James M. Buchanan, Center for Study of Public Choice, George Mason University $34.50

Economic Theory of Industry
MICHAEL WATERSON
"A fine little book that at last provides a comprehensive view of the new theory of industrial organization. Michael Waterson covers major topics in the theory of industry with a minimum of fuss and a maximum of clarity."—Richard J. Gilbert, Director, University of California at Berkeley Cloth $39.50 Paper $11.95

The Rules of the Game
The Logical Structure of Economic Theories
JOHANNES J. KLANT
Ina Swart, Translator
Starting from the Popperian notion that the scientific character of a theory depends on whether it is in principle "falsifiable" on the basis of empirical testing, this book shows that much economic theorizing is founded on untested and often untestable assumptions. The rules of the game, then, are those principles required if economists are to have a methodologically sound basis for their discipline. $39.50

William Jaffé's Essays on Walras
DONALD A. WALKER, Editor
"This is a beautifully produced book and a pleasure to read. It is a fitting memorial to William Jaffé and an excellent accompaniment to his translation of Walras' *Elements* and his monumental edition of the *Correspondence of Leon Walras and Related Papers*. Donald Walker, the editor, has written a short and useful introduction and translated 3 papers previously in French, but basically, as a good editor should, lets these 19 papers of Jaffé speak for themselves."—*The Economic Journal* $49.50

Cambridge University Press

Value and Distribution in Capitalist Economies
An Introduction to Sraffian Economies
LYNN MAINWARING

Beginning with a Sraffian reconstruction of Ricardo's theory of prices and distribution, the book goes on to a consideration of Sraffa himself and of those contemporary economists (Leontief and von Neumann) whose contributions may be regarded as supplementing and complementing Sraffa's. The author then develops an integrated model of price determination and economic growth in capitalist economies and extends the model to cover international trade and joint-in production systems. Cloth $37.50 Paper $12.95

The Politics of the World-Economy
The States, the Movements, and the Civilisations
IMMANUEL WALLERSTEIN

This new collection of essays treats the political manifestations of the world system, dealing with such highly topical issues as the role of the USA and USSR in the world-system, the relations of Third World states to the capitalist "core", and the potential for socialist or revolutionary change. Cloth $39.50 Paper $9.95

Why Has Japan "Succeeded"?
Western Technology and the Japanese Ethos
MICHIO MORISHIMA

"...stands out from the rest...because of Professor Morishima's exceptional ability as an economist and his intimate native knowledge of Japan...admirable for the range of its insights and the modesty of its conclusions..."—Jeremy Hardie, *The Times Literary Supplement* Cloth $29.95 Paper $9.95

The Philosophy of Economics
An Anthology
DANIEL M. HAUSMAN, Editor

While including the classic discussions by figures such as Mill, Marx, Weber, Veblen, Keynes and many others, this work also provides a sampling of work recently undertaken on this topic. It includes a comprehensive introduction by Professor Hausman and an extensive bibliography. Cloth $39.50 Paper $14.95

The US-USSR Grain Agreement
ROGER B. PORTER

This study of the 1975 negotiations illuminates the parameters and constraints faced by policy-makers in a democratic, pluralistic system. The author investigates the interaction between the formulation and implementation of a policy, suggesting that who is seen as responsible for a policy is often as important as its content. About $29.95

all prices subject to change

Order from your bookstore or call our Customer Service department 1-800-431-1580 (outside New York State and Canada). MasterCard or Visa accepted.

Cambridge University Press • 32 East 57th Street • New York, N.Y. 10022

New Books on

OPEC and

KOREA'S MARKET EXPANSION

OPEC IN A CHANGING WORLD ECONOMY

Abbas Alnasrawi

OPEC is not a cartel, argues Abbas Alnasrawi, but a pragmatic economic organization reflecting the country-specific aspirations of a group of nations having only one element—crude oil—in common. In OPEC IN A CHANGING WORLD ECONOMY, Alnasrawi provides an analytical context for studying the behavior of OPEC and its relations to the oil companies, consuming nations, and less-developed countries.

Alnasrawi examines the behavior and determination of crude oil prices prior to the 1970s, the changes under OPEC, the pattern of relationships between OPEC member nations and one another—and between OPEC and the rest of the world. Informed and objective—yet sure to be highly controversial—OPEC IN A CHANGING WORLD ECONOMY concludes with an assessment of the likely direction of OPEC policies and actions during the rest of the 1980s.

$22.50

KOREA'S COMPETITIVE EDGE
Managing Entry into World Markets

Yung Whee Rhee, Bruce Ross-Larson, and Gary Pursell

How did Korea manage to expand its exports from less than $100 million a year in the early 1960s to more than $25 billion a year in the early 1980s? To find out the underpinnings of Korea's competitive edge, the authors asked more than 100 major Korean exporters what had been important for them in institutional support, in technological development, and in marketing overseas.

The authors argue that there is a strong interaction between exporting and the effectiveness of a country's economic institutions. Their findings also show how Korea's selectivity in the acquisition of technology and in the marketing of products overseas has been an important part of its success in the world marketplace.

Published for the World Bank

$19.95

THE JOHNS HOPKINS UNIVERSITY PRESS

Baltimore, Maryland 21218

The Share Economy

Conquering Stagflation

Martin L. Weitzman

"This is one of the most exciting books in economics I have read in several years. Indeed, one has the suspicion that the message of this book may be the most profound intervention in unemployment policy of capitalist countries since Keynes's." –John E. Roemer

$15.00

Modeling Japanese-American Trade

A Study of Asymmetric Interdependence

Peter A. Petri

In rigorous, quantitative detail, this book examines the structure of trade between Japan and the United States, tracing the evolution of trade interdependence and the causes of its increasing intensity.

Harvard Economic Studies, 156

$20.00

Resources, Values, and Development

Amartya Sen

Sen examines foundational problems of development economics in depth, paying particular attention to institutional and motivational features that call for going beyond traditional economic approaches.

$29.50

The Intellectual and the Marketplace

Enlarged Edition

George J. Stigler

"Professor Stigler writes with a wit that is even more to be cherished than his erudition. Economics would never have been called the dismal science if there had been a nineteenth-century Stigler." –*The Economist*

$17.50 cloth; $7.95 paper

Blacks and Whites

Narrowing the Gap?

Reynolds Farley

What have American blacks gained from the civil rights struggle? Applying techniques of demographic measurement to data from Census Bureau surveys, Farley tackles this controversial question.

Social Trends in the United States

$19.50

available in paperback

How We Live

An Economic Perspective on Americans from Birth to Death

Victor R. Fuchs

"*How We Live* is a veritable handbook for reformers. Those concerned with policies affecting the family and the labor market especially cannot afford to ignore the information Fuchs assembles." –*The New Republic*

$7.95 paper

Harvard University Press

79 Garden Street, Cambridge, Massachusetts 02138

INDEX OF ECONOMIC ARTICLES

prepared under the auspices of

The Journal of Economic Literature
of the
American Economic Association

√ Each volume in the **Index** lists articles in major economic journals and in collective volumes published during a specific year.

√ Most of the **Index's** volumes also include articles of testimony from selected congressional hearings in government documents published during the year.

√ No other single reference source covers as many articles classified in economic categories as the **Index.**

√ The 1977 volume contains over 10,500 entries.

Currently available are:

Volume	Year Covered
XI	1969
XII	1970
XIII	1971
XIV	1972
XV	1973
XVI	1974
XVII	1975
XVIII	1976
XIX	1977
XX	1978
XXI	1979

an indispensable tool for...

ECONOMISTS
REFERENCE LIBRARIANS
RESEARCHERS
TEACHERS
STUDENTS
AUTHORS

Future volumes will be published regularly to keep the series as current as possible.

Price: $50.00 per volume (special 30% discount to AEA members)

Distributed by

RICHARD D. IRWIN, INC. Homewood, Illinois 60430

Prophets of Regulation

Thomas K. McCraw

Thomas McCraw unfolds the history of government regulation in America through the life stories of four dramatically powerful men; Charles Francis Adams, Louis D. Brandeis, James M. Landis and Alfred E. Kahn.

"... a seminal work in the study of American regulation." —Alan Brinkley

Belknap $20.00

Collected Papers

Volume 3: Individual Choice under Certainty and Uncertainty

Kenneth J. Arrow

The third volume of Kenneth Arrow's *Collected Papers* concerns the basic concept of rationality as it applies to an economic decision maker.

Belknap $25.00

Collected Papers

Volume 4: The Economics of Information

Kenneth J. Arrow

Volume 4 includes essays on the demand for information, the economic value of screening devices, the effect of incomplete information on the structure of organization, futures markets, insurance, and the allocation of risk.

Belknap $22.50

Growth, Distribution, and Prices

Stephen A. Marglin

What makes capitalist economies tick? This watershed treatise analyzes the answer in the neoclassical, neo-Marxian, and neo-Keynesian theoretical traditions.

Harvard Economic Studies, 155 $25.00

Summing Up

The Science of Reviewing Research

Richard J. Light and David B. Pillemer

"Astute and wonderfully lucid, *Summing Up* will be a treasure for inexperienced researchers and contains messages for many experienced ones as well. The book goes beyond techniques for boiling down multiple studies to a single summary "effect size," showing how qualitative information from the original studies improves interpretation."

—Lee J. Cronbach

$17.50 cloth; $7.95 paper

European Urbanization

1500–1800

Jan de Vries

"An impressive, intellectually rewarding achievement. It fills a major gap in the literature by combining judicious use of empirical material with elegant development of models."

—E.A. Wrigley,
London School of Economics

$28.50

available in paperback

Regulation and Its Reform

Stephen Breyer

"An excellent summary of regulatory practice and policy... It tells us how regulation works, when it fails, where it should be changed, and what to do to change it."

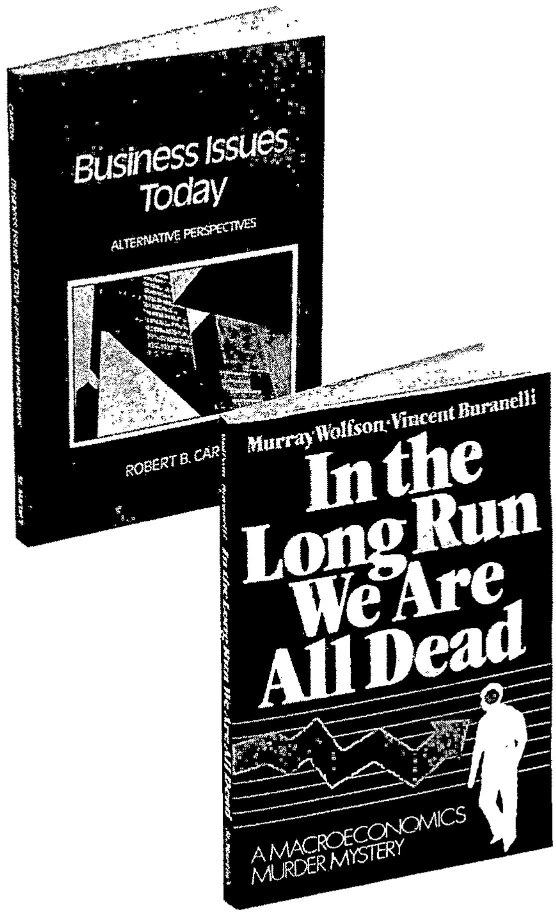
—*American Political Science Review*

$8.95 paper

Harvard University Press

79 Garden Street
Cambridge, MA 02138

Unusual—and unusually effective



Economics in Action
An Applications Book

BEVERLY J. CAMERON and NORMAN E. CAMERON, University of Manitoba

Designed to help students apply economic theory to real-life situations, *Economics in Action* offers a wide selection of newspaper and magazine articles (plus one book excerpt), advertisements, and cartoons dealing with key topics in economics. Questions that relate these source materials to economic concepts encourage students both to recognize theoretical concepts in practice and to analyze real behavior by using those concepts.

Paperbound. 192 pages. 1984
Also available in two separate volumes, *Microeconomics in Action* and *Macroeconomics in Action*
Instructor's Manual available

Business Issues Today
Alternative Perspectives

ROBERT B. CARSON, State University College, Oneonta, New York

Paperbound. 348 pages. 1984
Instructor's Manual available

"Very readable, challenging, and stimulating. I appreciate Carson's objectivity and unbiased approach in covering representative viewpoints on all sides of an issue, leaving conclusions to the students. I searched the market and this text met 99% of my expectations!"
—Robin E. Butler, Lakeshore Technical Institute

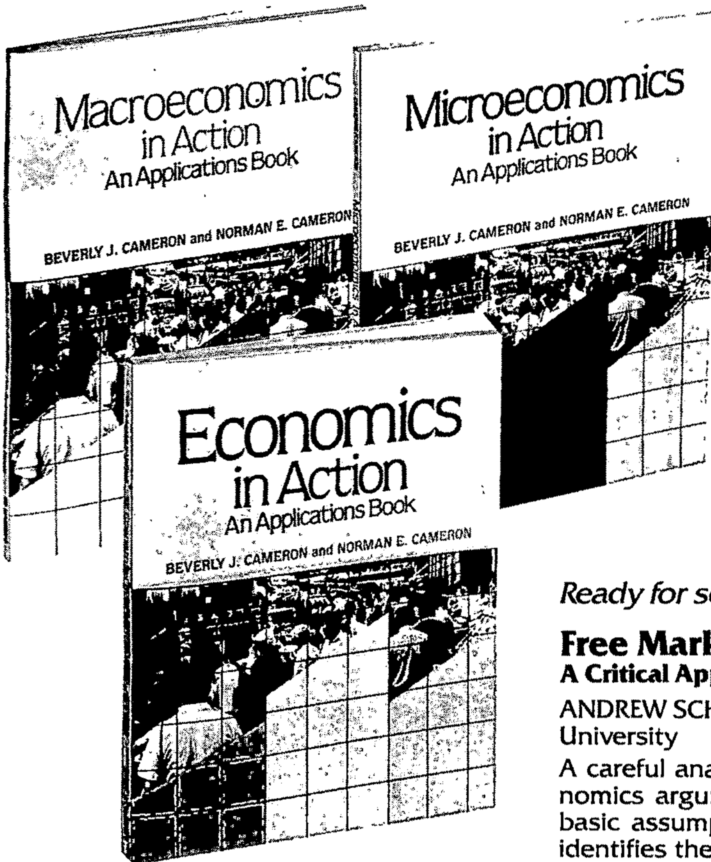
In the Long Run We Are All Dead
A Macroeconomics Murder Mystery

MURRAY WOLFSON, Oregon State University
VINCENT BURANELLI

Paperbound. 177 pages. 1984
Instructor's Manual available

"The authors have done a great job of using a murder mystery to teach macroeconomic theories. The text is well written and easy to understand. Every student of economics should read it. The anticipation of finding out 'who done it' coupled with the skillful way the theories are woven into the story line will hold the student's interest."
—Larry Olarewaju, Virginia State University

—supplements for your courses



The Contemporary International Economy

A Reader

Second Edition

JOHN ADAMS, University of Maryland, College Park

Intended as a supplement for courses in international economics and business, this book brings together 33 readings on the major issues in international economics today. The book provides a broad range of viewpoints, with selections by such noted economists as Martin Feldstein, Robert Baldwin, Herbert Grubel, Mordechai Kreinin, and Peter Drucker. The book is fully up to date: all but four of the readings are new to this edition.

Paperbound. 550 pages (probable)
Publication: February 1985

Ready for second semester—

Free Market Economics

A Critical Appraisal

ANDREW SCHOTTER, New York University

A careful analysis of the free-market economics argument, this book defines the basic assumptions of the argument and identifies their roots in the history of economic thought. Schotter reviews the major criticisms of the argument and raises some serious doubts about the blind advocacy of free market solutions to current social problems.

Paperbound. 148 pages. Publication: Just Published

To request a complimentary examination copy of any of these titles, please write us on your college letterhead specifying your course title, present text, and approximate enrollment. Address your request to:

ST. MARTIN'S PRESS Department JR
175 Fifth Avenue
New York, N.Y. 10010

New Titles For '85 From Prentice-Hall

FAIRMODEL STUDENT GUIDE: Economic Laboratory in Theory, Policy and Forecasting
Anthony Blackburn
Karl E. Case

MANAGERIAL ECONOMICS, Second Edition
Thomas J. Hailstones
John C. Rothwell

THE MAKING OF ECONOMIC SOCIETY, Seventh Edition
Robert L. Heilbroner

THE INTERNATIONAL ECONOMY
Peter B. Kenen

INPUT-OUTPUT ANALYSIS: Foundations and Extensions
Ronald E. Miller
Peter D. Blair

A GUIDEBOOK TO THE COMPARATIVE STUDY OF ECONOMIC SYSTEMS
Frederic L. Pryor

THE ECONOMICS OF INDUSTRIAL ORGANIZATION, Second Edition
William G. Shepherd

ECONOMICS OF THE FIRM, Fourth Edition
Arthur A. Thompson

MONEY, BANKING AND THE MACROECONOMY
Douglas Vickers

To request your examination copy(ies) for possible classroom adoption, write on your school letterhead to: Mr. Richard Hunter, Dept. J-393, Prentice-Hall, Inc., Englewood Cliffs, NJ 07632.

Prentice-Hall

NEOCLASSICAL POLITICAL ECONOMY
The Analysis of Rent-Seeking and DUP Activities
David C. Colander, Editor

Neoclassical political economics is a dynamic approach to understanding economic policy that integrates the effects of political action and policy into economic theory.

It is changing the way we perceive economic history, altering the way policy is formed, and redirecting the way economics is now and will be taught in the future.

Experts in the fields of politics, economic history, public choice, and international trade offer their perspectives on • historical and methodological approaches • the problems of modelling • new ways at looking at economic history • the application of empirical tests to theory.

Ride the new wave of political economics. Order NEOCLASSICAL POLITICAL ECONOMY today.
1984—ca 350 pp.—$32.00 cloth, ISBN 0-88410-999-2—LC 84-11124

INTERNATIONAL COMPARISONS OF PRODUCTIVITY AND CAUSES OF THE SLOWDOWN
John W. Kendrick, Editor
An American Enterprise Institute book

The most comprehensive treatment yet of international productivity comparisons. Includes a wealth of statistics on productivity trends and cycles.
1984—448 pages—$39.95 cloth, ISBN 0-88410-971-2—LC 84-3107

AMERICAN JOBS AND THE CHANGING INDUSTRIAL BASE
Eileen L. Collins and **Lucretia Dewey Tanner**, Editors
1984—240 pages—$32.00 cloth,
ISBN 0-88730-017-0—LC 84-14495

BANKING DEREGULATION AND THE NEW COMPETITION IN FINANCIAL SERVICES
Kerry Cooper and **Donald R. Fraser**
1984—210 pages—$32.00 cloth,
ISBN 0-88410-712-4—LC 84-6223

THE GLOBAL REPERCUSSIONS OF U.S. MONETARY AND FISCAL POLICY
Sylvia Ann Hewlett, Henry Kaufman, and **Peter B. Kenen**
A United Nations Association of the United States of America book
1984—256 pages—$25.00 cloth,
ISBN 0-88410-989-5—LC 84-14505

AFFORDABLE HOUSING
New Policies for the Housing and Mortgage Markets
Kenneth T. Rosen
A Twentieth Century Fund report
1984—ca 168 pages—$25.00 cloth,
ISBN 0-88730-005-7—LC 84-12419

NEW DIRECTIONS IN FEDERAL TAX POLICY FOR THE 1980s
Expanded Paperback Edition
Charls E. Walker and **Mark A. Bloomfield**
American Council for Capital Formation: Center for Policy Research
1984—464 pages—$16.95 expanded paperback edition, ISBN 0-88730-022-7—LC 84-18545

MONEY IN CRISIS
The Federal Reserve, the Economy, and Monetary Reform
Barry N. Siegel, Editor
Foreword by **Leland B. Yeager**
A Pacific Institute book
1984—392 pages—$35.00 cloth,
ISBN 0-88410-962-3—LC 84-9230

Please send me your free ECONOMICS CATALOG
NAME ________________________________

____

____________________ ZIP CODE ________

Mail to Mark Moquin, BALLINGER PUBLISHING CO.,
54 Church Street, Cambridge, MA 02138
or call (617) 492-0670

ECON1284

BALLINGER PUBLISHING COMPANY
Order Department
2350 Virginia Avenue, Hagerstown, MD 21740
To order, call toll free 1-800-638-3030

The real world
comes to the classroom . . .

Robert J. Barro

University of Chicago

MACROECONOMICS

Barro's Macroeconomics is the first modern presentation of the market clearing approach to macroeconomics. It provides an honest and realistic alternative to the cumbersome and unrewarding IS/LM approach of current intermediate texts.

> "Robert Barro has been a leading contributor to recent developments in the analysis of the overall behavior of the economy. He has now written a lucid, comprehensive, and authoritative exposition of the current state of our understanding of these phenomena. His text covers both the latest theoretical developments and the latest empirical evidence. It can be highly recommended on both levels."
>
> —Milton Friedman,
> Senior Research Fellow
> Hoover Institution on War, Revolution and Peace
> Stanford University

> ". . . What Barro does is not only build a macroeconomic model from the ground up but also one that is meant to explain observed economic phenomena . . . if only other textbook authors would try harder to do this, students would learn more and teachers would find their job much easier."
>
> —James Barth,
> The George Washington University

1984 580 pp.

You owe it to yourself and your students to examine this important new text. To request a complimentary copy, write to Lisa Berger, Dept. 5-1924. Please include course name, enrollment, and title of present text.

JOHN WILEY & SONS, Inc.
605 Third Avenue, New York, N.Y. 10158
In Canada 22 Worcester Road, Rexdale, Ontario M9W 1L1

5-1924

Keeping pace with the latest developments in economics

THE THEORY AND PRACTICE OF ECONOMETRICS
SECOND EDITION

George G. Judge, *University of Illinois*
William E. Griffiths, *University of Georgia*
Helmut Lütkepohl, *Universität Osnabrück*
Tsoung-Chao Lee, *University of Connecticut*

Now the most complete treatment of major econometric problems has been revised and updated. The authors present the most systematic and up-to-date view of econometric problems, their statistical consequences, remedies, alternatives, and future research. Includes new chapters on asympototic distribution theory, Bayesian inference, time series, and simultaneous equation statistical models. The distribution lag chapters have been rewritten to tie-in with time-series chapters.

Due January 1985 approx 1050 pp. ISBN 0-471-89530-X

Other New Titles:

AGRICULTURAL ECONOMICS AND AGRIBUSINESS, Third Edition
Gail L. Cramer, *Montana State University*
Clarence W. Jensen, *Montana State University*
January 1985 approx 475 pp. ISBN 0-471-87871-5

ECONOMICS OF PRODUCTION
Bruce R. Beattie, *Montana State University*
C. Robert Taylor, *Montana State University*
January 1985 approx 320 pp. ISBN 0-471-80810-5

1984 Titles of Interest:

AMERICAN MONEY AND BANKING
Maxwell J. Fry, *University of California*
Raburn M. Williams, *University of Hawaii*

MONETARY AND FINANCIAL ECONOMICS
James L. Pierce, *University of California, Berkeley*

URBAN LAND ECONOMICS
Michael A. Goldberg, *University of British Columbia*

CASES IN MANAGERIAL ECONOMICS, Second Edition
Bernard J. Winger, *University of Dayton*

To request a complimentary copy, write to Lisa Berger, Dept. 5-1925. Please include course name, enrollment, and title of present text.

JOHN WILEY & SONS, Inc.
605 Third Avenue, New York, N.Y. 10158
In Canada 22 Worcester Road, Rexdale, Ontario M9W 1L1

5-1925

Gain quick access to important information

The Food Manufacturing Industries
Structure, Strategies, Performance, and Policies

John M. Connor, Purdue University, Richard T. Rogers, University of Massachusetts, Bruce W. Marion, U.S. Department of Agriculture and University of Wisconsin, and Willard F. Mueller, University of Wisconsin
Foreword by A.C. Hoffman

The first study in two decades of the market structures, business conduct, and economic performance of the food, beverage, and tobacco manufacturing industries in the U.S.

The authors trace performance outcomes to the market structures of specific industries and conclude that more than one third exhibit poor economic performance. They support their position with detailed analyses of the role of large firms, advertising, mergers, foreign investment, long-term price stability, and food consumption patterns.

"The authors of this book are sensitive to the new dimensions of market power. . . . For years to come, their research and ideas will doubtless influence the views of both scholars and public policy makers."—from the Foreword

496 pages ISBN 0-669-08203-1 $38.00

The Politics and Economics of Organized Crime

Herbert E. Alexander and Gerald E. Caiden, University of Southern California

This book reveals the new face of organized crime in the 1980s and how it is affecting society. These experts revamp the old image of organized crime as an immutable institution of Western civilization. They find it today an institution capable of adapting to change. Essays consider the new legal and societal consequences of the growth and penetration of organized crime. The authors work toward developing a national strategy for combatting organized crime in this decade.

November ISBN 0-669-09342-4
ca. $20.00

Analyzing the Impact of Regulatory Change in Public Utilities

Michael A. Crew, Rutgers University, editor

This incisive and up-to-date analysis considers the rapid changes in regulation and regulated industries due to advances in technology, the AT&T divestiture, and the general emphasis on deregulation. Econometric studies on the cable TV industry and the cost structure of vertically integrated electric companies are included.

November ISBN 0-669-07341-5
ca. $27.50

Lexington Books
125 Spring Street
Lexington, MA 02173
(617) 862-6650 (212) 924-6460
Call our toll-free numbers
(800) 428-8071 or (800) 33-HEATH

Lexington Books

Please mention The American Economic Review *When Writing to Advertisers*

New Titles 1985 **From Harlan Davidson, Inc.**

Available in December '84

DYNAMICS OF BANKING

Edited by Thomas M. Havrilesky, *Duke University;* Robert Schweitzer, *University of Delaware;* and John T. Boorman, *International Monetary Fund*

A tidal wave of deregulation and innovation has swept over the financial services industry. It is provoking a revolution in bank management behavior, eroding the concept of distinct bank product and geographic markets, and compelling a realignment of supervisory and regulatory agencies. **Dynamics of Banking** satisfies the urgent need for a collection of readings that thoroughly covers the contemporary issues in this rapidly changing environment.

MODERN CONCEPTS IN MACROECONOMICS

Edited by Thomas M. Havrilesky, *Duke University*

Rampant growth in the size of government, runaway inflation, severe criticism of the rational expectations hypothesis, and a fundamental questioning of fiscal policy all create a demand for a collection of easily understood readings for macroeconomics and money and banking courses. **Modern Concepts in Macroeconomics** avoids complex mathematical and graphical formulations while it helps introduce students to the latest developments in macroeconomic theory and policy.

Of related interest:

INTRODUCTION TO MODERN MACROECONOMICS

By Thomas M. Havrilesky

A vigorous, flexible new text for the teaching of intermediate macroeconomics courses. Tables, charts, graphs, figures, glossary, index. 448 pages

CONTEMPORARY DEVELOPMENTS IN FINANCIAL INSTITUTIONS AND MARKETS

Ed. by Thomas M. Havrilesky and Robert Schweitzer 488 pages

MONEY SUPPLY, MONEY DEMAND, AND MACROECONOMIC MODELS, Second Edition

By Thomas M. Havrilesky and John T. Boorman 544 pages

CURRENT ISSUES IN MONETARY THEORY AND POLICY, Second Edition

Ed. by Thomas M. Havrilesky and John T. Boorman 512 pages

CURRENT PERSPECTIVES IN BANKING: Operations, Management, and Regulation, Second Edition

Ed. by Thomas M. Havrilesky and John T. Boorman 560 pages

Ten percent professional discount allowed. Examination copies for adoption consideration available if requested on departmental stationery, giving course number, estimated enrollment, and title currently in use.

HARLAN DAVIDSON, INC.
3110 North Arlington Heights Road
Arlington Heights, Illinois 60004

New dimensions of economics...

UNION POWER AND NEW YORK

Victor Gotbaum and District Council 37

Jewel Bellush and **Bernard Bellush**

This book is a detailed history of the New York City local, the largest and most successful municipal union in the city with 110,000 members.

The American Federation of State, County and Municipal Employees (AFSCME) is a public employee union which was organized in the mid-1930s.

ca. 368 pp. December 1984 ca. $33.95 ISBN 0-03-000122-6

WORLD BANKING AND FINANCE

Cooperation Versus Conflict

George Macesich

Examines the underlying forces of change in world banking and the on-going dialogue between developing-debtor countries and developed-creditor countries within the framework of the theory of cooperation.

ca. 176 pp. December 1984 $33.95
ISBN 0-03-001397-6

THE FEDERAL RESERVE SYSTEM

An Intentional Mystery

Thibaut de Saint Phalle

Explains in understandable terms the bank regulatory system in the United States.

ca. 304 pp. January 1985 ca. $35.95
ISBN 0-03-069289-X

Paper edition ca. $13.95 ISBN 0-03-069288-1

WITH CHARITY FOR ALL

Welfare and Society, Ancient Times to the Present

Merritt Ierley

"Fascinating. It should have considerable interest and certainly should be useful."
—Milton Friedman

An historical look at public provision for the poor.

236 pp. October 1984 $35.95
ISBN 0-03-000044-0

THE CHANGING STRUCTURE OF THE AMERICAN ECONOMY

Lessons From the Steel Industry

Zoltan J. Acs

ca. 208 pp. December 1984 ca. $26.95
ISBN 0-03-001092-6

Available through your local bookseller, or order directly from

PRAEGER PUBLISHERS 521 Fifth Avenue New York, N.Y. 10175

Please mention THE AMERICAN ECONOMIC REVIEW *When Writing to Advertisers*

American Journal of Sociology

Editors: Sherwin Rosen, Economics, University of Chicago
Christopher Winship, Sociology, Northwestern University

The *AJS* announces a special issue to explore and expand the dialogue between Sociology and Economics in areas of mutual interest. We seek substantively oriented papers that use analytical methods to further understanding of the systematic determinants of individual behavior and the analysis of social institutions. Areas in which we believe there is significant potential for fruitful interaction include: labor markets, sorting and selection, family, education, mobility decisions, occupational choice, lifecycle analysis, fertility and demography, and the structure of organizations. Proposals for exploration of yet other areas of mutual concern are welcomed. Papers that are substantively oriented and provide original structural analysis will be given highest priority in the selection process. However, papers that contrast approaches in a particular area or demonstrate the relevance and uses of work in one discipline from the other's viewpoint will be given serious consideration.

All papers will undergo the Journal's standard reviewing process and should be prepared according to the "Information for Contributors" in the *AJS*. We encourage potential contributors to submit a prospectus for possible papers, especially before initiating work specifically intended for the issue. Submit four copies of manuscripts. The relevant deadlines are listed below. Submissions should be sent to:

American Journal of Sociology
Special Issue
1130 East 59th Street
Chicago, Illinois 60637

Submission of Prospectus: February 1, 1985
Submission of Papers: June 1, 1985

JOB OPENINGS FOR ECONOMISTS

Available only to AEA members and institutions that agree to list their openings.

Annual Subscription Rates

U.S.A., Canada, and Mexico (first class):	$15.00, regular AEA members and institutions
	$ 7.50, junior members of AEA
All other countries (air mail):	$22.50, regular AEA members and institutions
	$15.00, junior members of AEA

Please begin my issues with:
☐ February ☐ April ☐ June ☐ August ☐ October ☐ December

Name_______________________________
First Middle Last

Address_______________________________

____
City State/Country Zip/PostalCode

Check one:
☐ I am a member of the American Economic Association.
☐ I would like to become a member. My application and payment are enclosed.
☐ (For institutions) We agree to list our vacancies in JOE.

Send payment (U.S. currency only) to:

THE AMERICAN ECONOMIC ASSOCIATION
1313 21st Avenue South
Nashville, Tennessee 37212

Econometric/Statistical Package

xSTAT™

Mainframe Capability, Accuracy and Speed for Your IBM Personal Computer

xSTAT: A Reliable, Economical Alternative to Mainframe Computing

- Is designed for the analysis of **VERY LARGE** as well as small data sets. It has no limit on sample size (total number of observations); and allows up to 50 variables per regression.
- Can generate and save an **X'X product matrix** the first time a data set is used. Subsequent regressions can be made from it to eliminate repetitive readings of input data normally required in statistical packages—very time-consuming when disk inputs are involved.
- Supports Intel 8087 80-bit Math Co-processor for **FAST NUMBER CRUNCHINGS** (has software emulation for systems without the co-processor to attain same accuracy). Double precision is used in matrix operations.
- Produces estimated coefficients that are identical to mainframe TSP results to four significant digits.
- Executes at **HIGH SPEED** when an X'X matrix is used to estimate regressions even for very large samples—it takes only seconds to complete a run.
- The output can be directed to disk, printer or screen. The display can be 132 or 80 columns wide and the statistics and coefficients can be printed in F-format (e.g., 123.456) or E-format (1.23456 + E02).
- Works in either batch or prompt mode; requires no special command language. **EASY TO USE.**

THE PACKAGE INCLUDES:

Data Reformatting and Transformation Program

- Converts formatted data into unformatted data and vice versa; aggregates over records; merges files; accepts data downloaded from mainframe or generated by other programs including Lotus 1-2-3, Visicalc, dBase II, WordStar, other word processors and editors.
- Allows selection of subsamples based on value range or sequential order. Converts fractions to decimal numbers (e.g., stock quotes, 32⅛ => 32.125).
- Has 14 mathematical/logical transformation functions for vector multiplication, division, addition, subtraction, logarithm, exponent, lagged variables, first difference, dummy variables, and time trend.

Descriptive Statistics

- Minimum, maximum, mean, standard deviation, sum of squares, correlation coefficients, **CROSS-TABS,** etc.

Ordinary Least Squares (Multiple regressions) and Two-stage Least Squares

- Generates variance/covariance matrix of coefficients, standard error, t-statistic, R-square, adjusted R-square, Durbin-Watson statistic and F-statistic. Calculates predicted values, residuals, and produces residual plot.
- Allows subsample selection; can also estimate regressions from product moment matrix (contingency table).

Pooling Cross-Section and Time-Series Data

- Estimates regressions for samples with multi-period observations (e.g., panel study). Automatic stacking of data matrix (diagonal block matrix). Estimates coefficients and F-statistics for testing various hypotheses of (1) a common intercept and slope; (2) a common slope and different intercepts; (3) a common intercept and different slopes; (4) different intercepts and different slopes.

SYSTEM REQUIREMENTS

IBM PERSONAL COMPUTER (PC or XT), COMPAQ or IBM PC compatible systems: 1 floppy disk drive or a hard disk system; DOS 1.1 or 2.0 and 192K memory.

PRICE: $250 for the complete package. (Accept checks, Visa & MasterCard. Or order through your local dealer.)

RELATED PRODUCT

Spreadsheet Forecasting Templates for Lotus 1-2-3: reads the estimated coefficients and inverse X'X matrix, saved optionally by xSTAT, directly into 1-2-3 for making "what if" forecasts and producing graphs. No keyboard data entry is needed. The templates are imbedded with formulae for estimating the standard error and confidence interval for the forecasted dependent variable that changes as the forecasted independent variables are varied. **PRICE $100**

MING TELECOMPUTING INC.

Telecommunications and Statistics for Microcomputers

P.O. BOX 101 Lincoln Center, MA 01773 (617) 259-0391

(SAMPLE OUTPUTS ARE AVAILABLE UPON REQUEST.)

OECDpublications

ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT

Main Economic Indicators: Historical Statistics 1964-1983

The new edition of this information-packed volume presents 20-year runs of indicators for such economic variables as GNP; production, deliveries, and stocks by industry; rate of capacity utilization; housing starts; wholesale and retail sales; employment and unemployment; producer and consumer prices; imports and exports; and much more. The base year used is 1980, and data is generally provided on an annual, quarterly, and monthly basis.

31 84 21 3 September 1984 656 pages ISBN 92-64-02616-9 $35.00

OECD Employment Outlook, 1984

OECD's latest forecasts regarding employment in OECD countries. The report finds that while employment growth is expected to pick up in most Member countries, it is not likely that unemployment will fall significantly outside the United States. Part One covers labor market developments and prospects and Part Two looks at such key topics as the contribution of services to employment, the importance of long-term job attachment, the effect of relative wages on youth unemployment, and the financial consequences of unemployment.

81 84 08 1 October 1984 118 pages ISBN 92-64-12621-X $13.00

Agricultural Trade with Developing Countries

Reviews trends in OECD agricultural trade with developing countries during the 1972/1980 period and examines the implications of these developments for policies in OECD Member countries.

51 84 02 1 June 1984 114 pages ISBN 92-64-12579-5 $11.00

New Forms of International Investment in Developing Countries

Examines how joint ventures, licensing, management- and production-sharing contracts, international subcontracting, and turnkey operations are affecting international investment in developing countries.

41 84 02 1 July 1984 140 pages ISBN 92-64-12590-6 $13.00

High Unemployment: A Challenge for Income Support Policies

Examines problems that the unemployment compensation systems of Member countries face during periods of high unemployment.

81 84 03 1 July 1984 254 pages ISBN 92-64-12561-2 $20.00

Exchange Rate Management and the Conduct of Monetary Policy

Examines the role of the exchange rate in policy formation, analyzing how the conduct of monetary policy affects exchange rates and how exchange rate management influences monetary developments. It also assesses the trade-off between domestic monetary objectives and exchange rate stability and considers its implications for policy.

11 84 01 1 October 1984 ISBN 92-64-12606-6 $18.00

Available from:

OECD Publications and Information Center
1750-E Pennsylvania Avenue, N.W.
Washington, D.C. 20006-4582 Tel.: (202) 724-1857

OECD
OCDE

ECONOMICS

22 JAN 1985

1985

NEW:

- Rosen
  PUBLIC FINANCE

- Blair & Kaserman
  ANTITRUST ECONOMICS

REVISIONS:

- Maurice & Smithson
  MANAGERIAL ECONOMICS, 2nd Edition

- Reynolds
  MACROECONOMICS: Analysis and Policy
  5th Edition
  MICROECONOMICS: Analysis and Policy
  5th Edition

- Tullock & McKenzie
  THE NEW WORLD OF ECONOMICS: Explorations into the Human Experience, 4th Edition

- Shepherd
  PUBLIC POLICIES TOWARD BUSINESS
  7th Edition

- Begin & Beal
  THE PRACTICE OF COLLECTIVE BARGAINING
  7th Edition

- Rowan
  READINGS IN LABOR ECONOMICS AND LABOR RELATIONS
  5th Edition

- Bornstein
  COMPARATIVE ECONOMIC SYSTEMS: Models and Cases
  5th Edition

1984

- Wonnacott
  MACROECONOMICS
  3rd Edition

- Seo
  MANAGERIAL ECONOMICS: Text, Problems, and Short Cases
  6th Edition

Examination copies for adoption consideration available upon request. Please indicate course title and text presently used.

Richard D. Irwin, Inc.
Homewood, IL 60430